

EARWITNESS CHARACTERISTICS AND SPEAKER IDENTIFICATION  
ACCURACY

By

GEA DEJONG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998

For Klaus

## ACKNOWLEDGMENTS

I want to express my deep gratitude to Dr. Harry Hollien, my chair. He has taught, questioned, encouraged, corrected, pushed, listened to, and guided me through the process of obtaining a doctorate of philosophy. However, his greatest gifts have been his constant demand for perfection and his endless enthusiasm for research.

I wish to thank my committee members, Dr. Anton Broeders, Dr. Samuel Brown, Dr. Carl Crandell, Dr. Anne Wyatt-Brown and Dr. Russell Bauer for all the time, effort, concern, and good times throughout my dissertation.

I want to express my gratitude to two people that have been my guardian angels during my stay: Dr. Jean Casagrande and Dr. Marie Nelson supported me in any way possible to ensure that I was able to continue my degree in their department. I thank them profusely.

I thank Dr. John Dickson for his statistical advice. Whenever I left his office, the world seemed such a logical place and the data smiled at me. There was also a statistically significant trend for Dr. Dickson to “forget” charging me for the consultation.

I thank my friend Robin Mukherjee for helping me out at a time that my research seemed one big question mark. He very patiently guided me through an analysis process that helped turning results into conclusions.

My official papers and documents would not have looked so elegant if I had not had the help of my landlady, Mrs. Ruth Duncan. Also, her concern for others and her cheerfulness and laughs made her the best housemate ever.

I would not have completed this degree without special friends who listened, shared, laughed, supported me, showed me their latest computer tricks and cheered me up with their E-mails.

I thank Reva Schwartz for being such a wonderful colleague, a knowledgeable and helpful co-worker and a great friend. She taught me all I wanted and did not want to know about American culture and thanks to her, I am finally able to produce English expressions and refrain from language that is considered improper (but is still used by everyone else in Gainesville....) I am happy to know that she successfully started your great “life-after-graduation”!

I thank Wayne King for the great chats and for helping me out when the wrong lights were blinking, the cables were gone, the lab trolls had misplaced crucial tapes and Mr. Tucker Davis decided to remain silent that day.

I thank my dear friend Satchi Venkataraman for the great times, the most exquisite Indian cooking and for the fresh milk in my refrigerator that was always there when I came back from long trips.

This dissertation is dedicated to Klaus Müller for his love, emotional and financial support, but most of all for his amazing patience. His two-year wait has impressed me and many others! He listened over and over again to my long stories of research frustration

and misery, even though they all started sounding similar. No words can express how I appreciate his love and support!

I wish to express the greatest thankfulness to my parents. First of all, they have given me a wonderful childhood, and have been supportive emotionally and financially, throughout my life. I am thankful for their encouragement and I share the honor of this degree with them. I also thank my brothers and sisters for their warmth and support; they have always made sure that coming home was a “feest.”

## TABLE OF CONTENTS

|   | <u>page</u> |
|---|-------------|
| ACKNOWLEDGMENTS .....   | iii         |
| LIST OF TABLES .....  | ix          |
| LIST OF FIGURES .....   | x           |
| ABSTRACT .....  | xi          |
| 1. REVIEW OF THE LITERATURE .....                                 | 1           |
| Introduction .....  | 1           |
| Speaker Recognition .....   | 2           |
| Approaches to Speaker Identification .....                        | 3           |
| Speaker Identification by Visual Inspection of Spectrograms ..... | 4           |
| Speaker Identification by Machine .....                           | 5           |
| Speaker Identification by Listening .....                         | 8           |
| Research on the Aural Perceptual Approach .....                   | 9           |
| General Research .....  | 9           |
| Listener's familiarity with the voice .....                       | 9           |
| Quality of the speech sample .....                                | 10          |
| Speaker's disguise .....  | 10          |
| Speaker's distortion .....  | 12          |
| Uniqueness of the speaker's voice .....                           | 12          |
| Non-contemporary speech samples .....                             | 13          |
| Length of speech sample .....                                     | 13          |
| Research Concerning the Listener .....                            | 14          |
| Listener's natural speaker identification ability .....           | 14          |
| General/forensic training of the listener .....                   | 14          |
| Gender of the listener .....                                      | 15          |
| Age of the listener .....   | 16          |
| The validity/reliability of the listener .....                    | 17          |
| Listener's familiarity with the language/dialect .....            | 17          |

|   |    |
|---|----|
| Research Concerning the Voice Lineup .....                      | 18 |
| Latency between first confrontation and identification .....    | 20 |
| Similarity of foils to target voice .....                       | 21 |
| The number of voices .....                                      | 21 |
| Earwitness' assumptions .....                                   | 22 |
| Earwitness' confidence .....                                    | 22 |
| Summary .....   | 23 |
| Aural Perceptual Issues Which Have Not Been Studied .....       | 23 |
| Memory of the Earwitness .....                                  | 24 |
| Research on earwitness identification and memory .....          | 24 |
| Research on memory in general .....                             | 25 |
| Auditory Skills of the Earwitness .....                         | 31 |
| Research on earwitness identification and auditory skills ..... | 31 |
| Research on auditory skills in general .....                    | 31 |
| Musicality of the Earwitness .....                              | 36 |
| Research on earwitness identification and musicality .....      | 36 |
| Research on musicality in general .....                         | 37 |
| Objectives of This Research .....                               | 41 |
| 2. METHOD .....   | 43 |
| The Subjects .....  | 44 |
| General Subject Selection .....                                 | 44 |
| Selection of LOW-SPID and HIGH-SPID Groups .....                | 45 |
| The Speakers and Speech Samples .....                           | 50 |
| Assessments of the Selected Subjects .....                      | 52 |
| Memory Assessment .....   | 52 |
| 1. Mental control .....   | 54 |
| 2. Logical memory I and II .....                                | 55 |
| 3. Verbal paired associates I and II .....                      | 55 |
| 4. Digit span .....   | 56 |
| 5. Auditory priming .....                                       | 58 |
| Psychoacoustic Assessment .....                                 | 61 |
| 1. Speech reception in noise test .....                         | 61 |
| 2. Frequency selectivity .....                                  | 63 |
| 3. Temporal resolution .....                                    | 66 |
| Musicality Assessment .....                                     | 67 |
| 1. Pitch discrimination .....                                   | 68 |
| 2. Intensity discrimination .....                               | 69 |
| 3. Rhythmic discrimination .....                                | 70 |
| 4. Timbre .....   | 70 |
| 5. Tonal Memory .....   | 71 |
| Pilot Study .....   | 72 |

|  |     |
|--|-----|
| 3. RESULTS .....   | 73  |
| Introduction .....   | 73  |
| Memory Assessment .....  | 74  |
| Psychoacoustic Assessment .....                                  | 83  |
| Musicality Assessment .....                                      | 87  |
| Summary of the Results .....                                     | 94  |
| Fitting a Model .....  | 95  |
| 4. DISCUSSION AND CONCLUSION .....                               | 104 |
| Introduction .....   | 104 |
| Discussion of the Results of the Assessment of Memory .....      | 104 |
| Discussion of the Results of the Psychoacoustic Assessment ..... | 110 |
| Discussion of the Results of the Assessment of Musicality .....  | 111 |
| Conclusion .....   | 114 |
| ABBREVIATIONS .....  | 120 |
| APPENDICES .....   | 121 |
| A MEDICAL QUESTIONNAIRE .....                                    | 121 |
| B VOICE LINEUP SENTENCES .....                                   | 124 |
| C UNIVARIATE SAS PLOTS .....                                     | 125 |
| D CORRELATION ANALYSIS .....                                     | 142 |
| E INTERACTION ANALYSIS .....                                     | 144 |
| F ESTIMATES AND P-VALUES OF THE SECOND MODEL .....               | 145 |
| LIST OF REFERENCES .....   | 146 |
| BIOGRAPHICAL SKETCH .....  | 174 |



## LIST OF TABLES

| <u>Table</u>   | <u>page</u> |
|--|-------------|
| 1. Mock witness test results .....   | 51          |
| 2. Overview of tests .....   | 53          |
| 3. Means and standard deviations for the memory tests .....                        | 75          |
| 4. Results of the Two-Sample One-Tail T-Test performed on the memory data .....    | 82          |
| 5. Means and standard deviations for the psychoacoustic tests .....                | 84          |
| 6. Results of the Two-Sample One-Tail T-Test performed on the psychoacoustic data. | 87          |
| 7. Means and standard deviations for the music tests .....                         | 88          |
| 8. Results of the Two-Sample One-Tail T-Test performed on the music data .....     | 93          |
| 9. Tests that satisfy the requirement for entering the model .....                 | 97          |
| 10. Estimates and p-values of the first model .....                                | 99          |
| 11. Estimates and p-values of the final model .....                                | 101         |

## LIST OF FIGURES

| <u>Figure</u>  | <u>page</u> |
|--|-------------|
| 1. Histogram: SPID Scores and their frequency . . . . .                          | 48          |
| 2. Descriptive univariate SAS-plot of the digit span backward data . . . . .     | 78          |
| 3. Descriptive univariate SAS-plot of the attention/concentration data . . . . . | 80          |
| 4. Descriptive univariate SAS-plot of the MRT data. . . . .                      | 86          |
| 5. Descriptive univariate SAS-plot of the pitch data . . . . .                   | 90          |
| 6. Descriptive univariate SAS-plot of the tonal memory data . . . . .            | 92          |
| 7. A speaker identification model for assessing earwitnesses. . . . .            | 103         |

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

EARWITNESS CHARACTERISTICS AND SPEAKER IDENTIFICATION  
ACCURACY

By

Gea DeJong

May 1998

Chairperson: Professor Harry Hollien, Ph.D.  
Major Department: Linguistics

The earwitness lineup, also called a voice lineup, is a process by which a witness hears a series of voices and is asked to identify (if possible) one of the speakers. Lineups are employed in criminal investigations on an international basis, and the results are accepted in many courts. This study was designed to investigate the effect of earwitness characteristics on speaker identification accuracy. Experiments have shown that people exhibit a wide range in natural identification skills. Some individuals are quite good at this type of identification, even without any training, while others show poor-to-modest performances. Not much is known about these relationships; that is, those between the individual features of an earwitness and his/her success in identifying a speaker. This study, therefore, is an attempt to provide data that allow a better understanding of this

relationship. The focus was on the memory, auditory and musical skills of the subjects who exhibited either good or poor identification abilities. The experimental question was addressed: 1) do memory, auditory and musical skills of the earwitness affect the ability to identify speakers? and if so, 2) how do those characteristics influence accuracy? A group of 112 young women between 18 and 35 years volunteered for the study and were subjected to a speaker identification experiment. Subsequently, two groups were selected: they consisted of 14 women that scored highest on this task (designated the HIGH-SPID group) and the 13 with the lowest score (the LOW-SPID group). Memory, auditory, and musical skills were assessed of each individual in both groups, and the results compared by group. Statistical tests (two sample one-tail t-tests and logistic regression analysis) were employed; they demonstrated where the groups differed from each other and therefore which characteristics significantly affect speaker identification accuracy and which do not.

It appeared that factors that require high level cognitive processing are better predictors of an earwitness' ability to identify speakers than those that are associated with basic mental skills. Therefore, earwitnesses do not need to excel in the basic auditory and memory skills. It was also observed that listeners that exhibit a high degree of musical aptitude can be expected to perform well. In addition, the study showed that differences in intonation seem to be important cues for identifying speakers for earwitnesses involved in a voice lineup.

## CHAPTER 1 REVIEW OF THE LITERATURE

### Introduction

Professionals in the area of phonetics are interested in all facets of speech, for example, acoustics, physiology, perception as well as the characteristics and the classifications of speech and speech sounds. Forensic phonetics is a subdivision of phonetics. Its focus is primarily on jurisprudence and law enforcement. Currently, areas include speaker identification, speaker analysis, speech transcription, tape authentication, speech enhancement, and those involving the effects of psychological or chemical states (for example, intoxication or stress) on speech. Speaker identification is one of the most important tasks of the forensic phonetician; here he/she seeks to determine whether the voice of a given individual, an unknown, matches that of known talkers. Three approaches to identifying the speaker have been developed: identification by 1) listening to the voice, 2) using a machine, and 3) (in the past anyway) visual inspection of spectrograms. This research will deal with speaker identification by listening only. Here, different types of listeners can be used. For example, forensic phoneticians are employed when a voice recording of the criminal and suspects exists, but earwitnesses are unavailable. When there is no recording, but a lay earwitness is available, this person may carry out the identification. Although a substantial amount of research has been

conducted during the past 40 years, many issues in this area are as yet unexamined. One such area is the relationship between characteristics of earwitnesses and their speaker identification accuracy. Is it possible that certain listeners are relatively good at this type of identification, while others are not? If so, why? It is possible that differences in individual memory affect the ability to identify speakers or perhaps there are differences in auditory skills or musical abilities. These issues seem crucial to the speaker identification task. They have not yet, however, received research attention. Therefore, speaker identification accuracy will be studied as a function of memory, auditory, and music skills of the individual earwitness.

### Speaker Recognition

Speaker identification constitutes one of the two subareas of speaker recognition. Basically, speaker recognition is divided into speaker identification and speaker verification.

Speaker verification is concerned with validating the claim of an individual that he/she is, indeed, the target speaker. The verification procedure is usually generated at the request of the speaker, who wishes to be recognized. Here, speech samples of both the "unknown" and the known candidate are compared and a conclusion is reached. Speaker verification can be used to screen people requesting access to secure areas or to certain electronic systems (e.g., banking applications).

The speaker identification paradigm also involves the matching of known and unknown voices. In this case, a sample of an unknown voice -- often the voice of a

criminal -- is compared to a set of samples produced by known speakers. Here, the speaker prefers not to be recognized. The identification process is in most cases initiated at the request of individuals other than the speaker him/herself, e.g. law enforcement officers or attorneys.

The success of any speaker recognition procedure depends on selecting features for analysis that result in the interspeaker variability being greater than the intraspeaker variability. Interspeaker variability can be defined as the difference in the speech characteristics of two (or more) different speakers. The intraspeaker variability is the variability which can be observed if a given individual produces the exact same utterance twice. In other words, the features of investigation, either by machine or by listening, should 1) only discriminate between speech segments coming from different speakers and 2) show a very small variability between segments uttered by the same speaker. Speaker identification is a more complex process than speaker verification as in speaker verification the set is of limited size, the speakers are cooperative, the text is sometimes pre-defined, and the recording of high-fidelity. In speaker identification, however, the set of speakers involved is much larger, speakers are uncooperative, the text is freely chosen, and the speech recordings often are poor.

#### Approaches to Speaker Identification

As stated earlier, three different methods have been developed in an attempt to carry out speaker identification; they are 1) aural-perceptual approaches, 2) visual

inspection of spectrograms and 3) processing by machine. Because the aural perceptual method, identification by listening, is the focus here, it will be discussed last.

### Speaker Identification by Visual Inspection of Spectrograms

Gray and Kopp (1944), Potter (1945) and Potter, Kopp and Green (1947) were the first investigators to propose the use of a certain type of spectrograms in speaker identification. A spectrogram of this type is a representation of a signal in a frequency-by-time-by-amplitude dimension. Over time, it displays those frequencies at which energy is concentrated, and (generally) how much energy is present. It was not until the early sixties, however, that visual inspection of spectrograms became popular. At that time, Lawrence Kersta, an engineer from Bell Telephone Laboratories, began to promote this method as a tool for law enforcement. In his 1962 paper, he claimed that individual spectrograms can be regarded as being as valid as fingerprints, and that they are distinctive enough to identify an individual. Using this analogy, he called those spectrograms "voiceprints." His statement raised the curiosity of many forensic researchers and considerable research was done in the late sixties and seventies investigating the use of this technique. Even though most practitioners hoped that research would show that it was useful for forensic research, many were concerned about the unacceptably high rates of incorrect speaker identification found in the majority of experiments. That is, nearly all researchers found very high error rates whereas Kersta (1962a, 1962b) reported close to perfect scores -- i.e. only 1% incorrect identification. He therefore claimed the voiceprint method to be a valid one.



By the late sixties and early seventies, most forensic phoneticians (Bolt et al., 1970, 1973; Hollien, 1971, 1977; Hollien and McGlone, 1976; Ladefoged and Vanderslice, 1967) had rejected the method as a reliable tool for law enforcement. The majority of the scientific community could not accept the method. Too many discrepancies existed in research and unacceptably high error rates were obtained in experiments, particularly in those approaching real-world conditions (Hecker, 1971; Hollien, 1990). The voiceprint method, however, is still practiced by some law enforcement agencies; ones such as the FBI (Koenig, 1986). It was also discussed again by Thomas Owen at the American Academy for Forensic Scientists Conference in 1996, Nashville, Tennessee, where he stated that as a result of modern methods and adherence to standards, voice identification (including voiceprints) is a very effective tool in both the law enforcement community and for security/access control. However, he did not support this contention with any evidence -- scientific or otherwise.

### Speaker Identification By Machine

Another approach to speaker identification is by machine or computer processing. A generic automatic speaker recognition system consists of a pre-processor extracting the features from the signal, a matching module controlling the comparison process and a database with speaker references. Both identification and verification systems can be either text independent or text dependent. In a text independent system, arbitrary speech is used in the comparisons. In a text dependent system a particular string of phonemes or words is established as a basis of the identification/verification process. In the field of

speaker recognition, most automatic recognition systems are designed for use in verification as their performance has shown to be high enough for commercial exploitation. Some examples of speaker identification systems, however, are SAUSI (Hollien, 1990), a semiautomatic system, the probabilistic system reported by Wolf et al. (1983), the vector quantization approach by Soong et al. (1985) and the identification system of Webb et al. (1993); this latter method is based on Hidden Markov Models. Recently, neural networks have been gaining popularity for speaker identification use purposes (Bennani and Gallinari, 1995). Even though speaker identification systems show correct identification scores that are slightly lower than verification systems, they are useful tools when used in addition to aural perceptual judgement (Hammersley and Read, 1996). The main problem researchers in this area have to face is the same as for any identification procedure: the variability of the signal within the speaker. As stated, the performance of the system depends to a large extent on the selection of those features that minimize the intra-speaker variability while maximizing the inter-speaker variability. The effectiveness of various features has been studied extensively (Hollien, 1990; Sambur, 1975; Rosenberg, 1976). Speaking Fundamental Frequency (also F0), for example, has been suggested and used in several voice comparison algorithms (Atal, 1972; Castellano et al, 1997; Compton, 1963; Doddington, 1976; Hollien et al., 1975; Houde, 1995; Iles, 1972; Jiang, 1996; LaRiviere, 1971; Lumnis, 1972; Markel et al., 1977; Mead, 1974; Nakasone and Melvin, 1988; Nolan, 1983; Rosenberg and Sambur, 1975; Wohlford, 1980). Even though F0 has been found to be a reliable feature, it has several disadvantages. F0 is sometimes difficult to measure in noisy environments, it may be

unstable over time, and it is easy to mimic. Another robust feature is the Long-term-speech-spectrum (Bricker et al., 1971; Clarke and Becker, 1969; Doddington, 1970; Furui, 1978; Kosiel, 1973; Majewski and Hollien, 1974; Zalewski et al., 1975). The LTS-vector has shown high accuracy levels, and it is also resistant to the effects of speaker stress and to limited passband conditions. Identification accuracy decreases, however, when a speaker disguises the voice (Doherty, 1976; Doherty and Hollien, 1978; Hollien, 1990; Majewski and Hollien, 1974; Zalewski et al., 1975). Speech intensity also has been shown to be an effective measure (Doddington, 1971; Lumnis, 1973; Rosenberg and Sambur, 1975). As with F0, intensity has the disadvantage that it is not difficult to mimic or to manipulate. Second, this factor contributes only in a minor degree to the identification process if field data are used (Hollien et al., 1990a; Hollien et al., 1990b). Formant frequencies also could distinguish speakers but they are difficult to measure. Calculation methods of the formant frequencies, however, have been proposed by several authors (Calinski et al., 1970; Hollien, 1990; Jiang, 1995, 1996; McCandless, 1974; Meltzer and Lehist, 1972; Rajasekaran, 1984). Other features that have been investigated are derivatives of the signal- like filter bank magnitudes, Linear Prediction Coding spectral and cepstral coefficients, coarticulation characteristics, and timing/duration features (Atal, 1974; Carey et al, 1997; Doddington, 1974; Doherty, 1976; Doherty and Hollien, 1978; Furui, 1981; Jiang, 1995; Johnson et al., 1984; Luck, 1969; Rosenberg, 1976; Sambur, 1976; Su et al., 1974; Velius, 1988; Wohlford, 1980) In general, the most reliable features appear to be fundamental frequency and long-term-spectra (Naik, 1990). Other features may be too complex to measure or, presently, do not show a high enough reliability for speaker identification.

### Speaker Identification by Listening

Do people remember anything about a voice when they have listened to someone speaking? Researchers in the area of speech perception found that during word recognition, voice information is not discarded, but is represented in the long-term memory representations of spoken words (Goldinger, 1996; Meehan and Pilotti, 1996; Palmeri et al., 1993; Sheffert and Fowler, 1995). Voice-specific memory can even serve as a retrieval cue for word recognition (Sheffert and Fowler, 1995). Pisoni (1993) investigated the storing of non-linguistic information. He found that listeners apparently retain information in long-term memory about the speaker's gender, dialect, speaking rate, and emotional state. These are attributes of speech signals that are not traditionally considered part of phonetic or lexical representations of words. The fact that listeners remember voice-related characteristics explains why the third approach, the aural-perceptual procedure, is possible at all. "Listening" of course is the oldest of the three identification approaches as it does not depend on any device external to the auditor. It is known that this process has been used in court cases dating back to the 17th century (Hollien, 1990). More recently, aural perceptual identification testimony has been accepted as legal evidence by courts all over the world. For example, in the state of Florida, the admittance of this type of evidence was documented as early as 1907 (Mack vs. State of Florida). Identification of this type is carried out on a regular basis in normal daily life as part of human interactions. When it involves the judicial process, it may be done by a forensic phonetician or an earwitness. Substantial research has been carried out on aural perceptual speaker identification.

### Research on the Aural Perceptual Approach

One of the earliest studies on perceptual speaker identification was carried out by McGehee (1937). Her efforts were in response to the kidnaping case of the Lindbergh child (State vs. Hauptmann, 1935). Here, the court allowed Col. Lindbergh to make a voice identification in court, two years after he heard it. McGehee was concerned about the effect of the time delay of two years on the reliability of the identification. In her study, she had listeners identify speakers after different time intervals (i.e. 1, 2, 4, and 20 weeks). The results clearly indicated that the identification procedures used at that time could not be justified: time appeared to have a detrimental effect on identification accuracy. A substantial amount of research has been carried out over the years since that time and much more is known about the factors affecting identification accuracy. For example, factors like familiarity with the voice, the quality of the speech sample and the latency between first confrontation with the voice and the identification are found to be of great importance to the aural perceptual process.

### General Research

Listener's familiarity with the voice. Speakers who are known to the listener have been found to be easier to identify than are unfamiliar speakers (Abberton and Fourcin, 1978; Hecker, 1971; Hollien and Thompson, 1990; Hollien et al., 1982; Pollack et al., 1954; Rose and Duncan, 1995). For example, Hollien et al. (1982) showed that auditors who listen to the speech of individuals with whom they are very familiar can be expected to identify them at very high levels of accuracy even for conditions where the sample was

produced while the speaker was experiencing mild stress (98%). Individuals who do not know the speakers can be expected to be able to learn to identify them quickly at levels well above chance (40%), but still at levels that are substantially lower than for familiar speakers.

Quality of the speech sample. For a forensic phonetician or an earwitness, working with a speech sample that was recorded with high-fidelity equipment and that contains no interfering background noise, facilitates the identification process significantly. Unfortunately, tapes with low-quality recordings are very common. Steady-state noises (e.g. a 60Hz. hum) or intermittent noise like the closing of a door can be damaging to the speech signal. However, the interference coming from other speakers or music is more difficult for a forensic phonetician to deal with: this noise usually has its energy in the same frequencies as the targeted speech (Hollien 1990, 1992). A very common problem is the degradation of the signal, because it was recorded over a telephone line. Telephone transmission can severely damage the signal: it bandpass-filters the signal between 300Hz. and 3,400Hz.. As a result, recognition rates tend to drop slightly (Künzel, 1990, 1994; Rothman, 1977). In general, the quality of the speech material is an important factor in identification.

Speaker's disguise. One of the tools available to the criminal when communicating with the victim is to disguise his/her voice. Different types of disguise can be distinguished: one, where the speaker decides to alter the manner of vocal fold vibration such as with a

whisper, or by changing fundamental frequency for example. Other types occur at higher levels within the speech process. For example, one could mimic a dialect or language, simulate some kind of pathology, manipulate the vocal tract (lip protrusion, closing the nose) or even imitate an ungrammatical sentence structure. External objects or devices also can be used: included are electronic devices (Masthoff, 1996), or holding a pencil between the teeth while speaking (deFigueiredo and deSouza Britto, 1996). Listener's performance is usually severely degraded by the speaker's use of disguise, even though it still can be greater than chance (Carbonell et al., 1965; Endress et al., 1971; Hecker et al., 1968; Hirson and Duckworth, 1995; Hollien et al., 1982; McGehee, 1937; McGlone et al., 1977; Reich and Duke, 1979; Simonov and Frolov, 1973; Tate, 1978; Williams and Stevens, 1972). The disguise preference and the particular effects of the disguise on the signal have been studied by several researchers (deFigueiredo and deSouza Britto, 1996; Gfroerer, 1994; Hirson and Duckworth, 1995; Masthoff, 1996; McClelland, 1994; Reich and Duke, 1979). Reich and Duke (1979) studied six different "voice mode disguises", including normal aged, hoarse, hypernasal, slow, and free disguise. They found that the nasal and free disguise were the most effective modes of disguise in reducing performance (59.4% and 61.3%, respectively), whereas the undisguised condition led to significantly higher accuracy (92.3%) over any of the other conditions. Gfroerer (1994) and Masthoff (1996) both found a preference for the alteration of phonation (e.g., whisper, changed pitch, etc.). Masthoff noted that raising F0 was attempted only by males and lowering F0 only by females. To conclude, voice disguise is a powerful tool for a speaker that does not want to be recognized, since it severely degrades a listener's performance.

Speaker's distortion. It is not only disguise that can change a voice. Other conditions that are to a lesser degree under the direct control of the speaker can also affect the voice, like stress or anxiety, temporary health conditions like a cold, ingested drugs or alcohol (Constanzo et al., 1969; Fairbanks and Hoaglin, 1941; Fairbanks and Pronovost, 1939; Friedhof et al., 1964; Hicks, 1979; Hollien, 1990; Hollien and Martin, 1996; Naik, 1990; Scherer, 1974, 1977, 1979a, 1979b, 1981, 1986; Silverman and Silverman, 1975; Williams and Stevens, 1972). Even though, the cited conditions like a cold, for example, may not occur very often, the possibility of a speech signal affected by them should be taken into account in a forensic investigation.

Uniqueness of the speaker's voice. Certain voices are easier to recognize than others (Köster, 1981). In Papcun et al. (1989) listeners (n=90) had to recognize unfamiliar voices one, two and four weeks after the first exposure. They found that the number of correct identifications was about the same for each voice, but "hard-to-remember" voices were more often misidentified as the target voice than "easy-to-remember" ones. Their theory is that of the "prototype": voices (i.e., those which can be remembered as a prototype with certain extra features) do not decay as fast as voices which do not fit a prototype. They state that various results indicate that prototypes have a special status in memory. Bartlett (1932) suggested that forgetting tends to affect peripheral information more than abstracted prototypical information. With controlled experiments using artificial stimuli, Posner and Keele (1970) found that performance on prototypes decayed more slowly than performance on non-prototypes. From the research cited above, it can



be concluded that the type of voice, whether hard or easy to remember, affects the validity of a witness's judgement.

Non-contemporary speech samples. Samples of the same speaker recorded at different points in the speaker's life are called non-contemporary samples. Since a speaker's characteristics may change over time (Endress et al., 1971), it has been suggested that non-contemporary samples will make the speaker identification more difficult (Rothman, 1977). However, Hollien and Schwartz (1997) and Schwartz (1995) tested subjects with non-contemporary samples of different time-intervals ranging from a few weeks (4wk, 8wk, 32wk) to a period of years (6 y., 20 y). They found that up to the first 6 years the effect was only moderate, 15-20% identification error, but that after 20 years this went up to 67% error. Therefore, it can be concluded that voice characteristics seem to be quite stable over time. This means that the problem of voice-specific features that have changed over time only exists for a minority of cases -- those with a latency over six years.

Length of the speech sample. Pollack et al. (1954) were among the first to claim that one of the most effective factors for speaker identification was the duration of the signal. However, this is true only if it admits a larger statistical sampling of the speaker's speech repertoire (Bricker and Pruzansky, 1966). For longer periods of time there is no improvement or only moderate improvement (Compton, 1963; Cort and Murry, 1972; Pollack et al., 1954).

### Research Concerning the Listener

Listener's natural speaker identification ability. Many researchers in this field have noted the fact that their listener subjects showed a wide range of speaker identification skills (Bartholomeus, 1973; Bull et al., 1983; Coleman, 1973; Hollien and Köster, 1996; Hollien and Thompson, 1990; Hollien et al., 1995; Iles, 1972; Köster, 1981; Künzel, 1990, 1994; Stevens et al., 1968; Thompson, 1985b). The extremes can be extensive from one subject scoring close to 0% correct to another who almost seems to have a special developed sense for correctly identifying speakers. For example, in one group in the Hollien et al.'s (1982) study, subjects' percentage correct ranged from 0%-100%. Many phoneticians see this variability as being of major concern (Hollien et al., 1995; Stevens et al., 1968). First, if the auditor happens to be an individual at the lower end of the scale, then the validity of his/her judgement could be questioned. Thus, it would appear that a standardized test has to be developed to measure a listener's identification skills. The present study is designed to investigate these very relationships, i.e. those between earwitness characteristics and the ability to identify speakers.

General and forensic training of the listener. It has been shown that general training in phonetics increases identification accuracy only slightly, but specific training in forensic phonetics improves it considerably (Hirson and Duckworth, 1995; Hollien and Thompson, 1990; Huntley, 1992; Köster, 1981; Nerbonne, 1968; Shirt, 1984). For example, Köster (1981) found that, in the three experiments he carried out, the errors for the nonprofessionals ranged from 0%-33%. However, in no instance did his phonetician

auditors make an error. Shirt (1984) found only a slight improvement for the listeners with training in phonetics. However, the fact that the difference was so small, was mainly due to the extremely good performance of one lay auditor. In general, individuals with training can be expected to outperform listeners without an education in the field of phonetics.

Gender of the listener. Results about the relationship between gender and speaker identification accuracy do not seem to be conclusive. McGehee's extensive study (1937) with in total 554 male and 186 female listeners suggests that men's performance will be better than women's. However, Bull and Clifford (1984) state that, in their studies, female listeners perform more accurately than did males, and Thompson (1985a) with 240 subjects did not report any effects for sex of subject nor a sex of subject by sex of speaker interaction.

Can listeners identify the gender of the speaker? Research examining this question appears somewhat more conclusive than the investigations studying the former issue. Identifying the gender of the speaker can be done with a high level of accuracy (Coleman and Lass, 1981; Ingemann, 1968; Schwartz, 1968) even if the speech segments are only voiced fricatives. Ingemann (1968), who studied this type of fricatives, found that as the portion of the vocal tract in front of the constriction increases, so does the identification accuracy of the speaker's sex. For example, the highest accuracy was obtained by using the [h], which involves the whole vocal tract. In short, research on gender differences in speaker identification is inconclusive, but it has shown that identification of the speaker's gender is relatively easy.

Age of the listener. Investigators have noted a poorer identification performance in children and elders (Bartholomeus, 1973; Bull and Clifford, 1984; Clifford, 1980a; Künzel, 1990). A moderate level of identification seems to be possible already at the age of a few months (Mehler et al., 1978; Friedlander, 1970). However, by age 10, children may reach adult levels of correct identification (Mann et al., 1979). If one considers speaker identification as a form of pattern recognition, with the patterns being auditory, it also may be interesting to look at some studies in psychology, especially one by Gibson and Gibson (1955). It seems that younger children tend to overgeneralize in pattern recognition. The Gibsons often required children in their experiments to select those patterns from a set which exactly matched a standard. A nonsense form on a card is shown for five seconds. Next, the subject is shown a series of cards with pictures with some pictures exactly matching the target form and the rest differing from it (e.g., difference in number of coils, horizontal stretching or compression, or right-left reversal). Their results show that children between six and eight years old identify nearly all the items as matching the standard. Adults, however, rarely find that the undifferentiated items match. The results of a group of older children were in between those extremes. In general, research suggests that in law enforcement, one should be careful when the earwitness is a child or an older adult.

The validity/reliability of the listener. Very recently, researchers have shown interest in developing a procedure to measure the validity and the reliability of listeners (Broeders and Rietveld, 1995; Künzel, 1990; Huntley and Pass, 1995). Validity refers to the

judgement being correct; reliability to the consistency of the responses. To test validity, some phoneticians have suggested the running of multiple trials with the same target voice (Künzel, 1990; Huntley and Pass, 1995). For example, a standard practice at the German Bundeskriminalamt is a repeated confrontation, after which the reliability value is calculated from all listener scores. It is assumed that high reliability correlates with a high response validity. In an attempt to find an appropriate validity test, Huntley and Pass (1995) gave their listeners a paired comparison test before the actual voice lineup. In the pretest, the subjects were required to listen to pairs of speakers and indicate whether they were the same or not. However, the results were somewhat discouraging: no correlation was found between the score of the paired comparison test and the score of the voice lineup. It is clear that more research is necessary to develop tests that assess the validity and reliability of the witness.

Listener's familiarity with the language/dialect. Speaker identification with a speaker who does not speak the listener's language or accent probably will result in a lowered accuracy (Goggin et al., 1991; Hollien et al., 1982; Köster et al., 1995; Schiller and Köster, 1996; Thompson, 1987). Thompson (1987) found that monolingual English listeners identified English speakers significantly better than they did either Spanish speakers or English speakers with a Spanish accent. Goggin et al. (1991) stated that voice identification is increased approximately twofold when the listener understands the language compared to when the speech sample is in a foreign language. Therefore, extra care should be exhibited when a phonetician or earwitness is not familiar with the speaker's language or dialect.

### Research Concerning the Voice Lineup

The findings cited above all apply to aural perceptual research in general -- it concerns any investigation where speaker identification is involved. The listener could be, for example, a forensic phonetician, but also a lay person like a victim or earwitness. The focus of this research, however, is speaker identification by earwitnesses only. The police will search for that type of listener, when there is no recording of the criminal's voice. There are two possible identification formats, the single versus the multiple confrontation (Broeders and Rietveld, 1995). In the case of the single confrontation, the witness is exposed to the voice of the suspect only. In the latter case, the witness is presented not only with the suspect's voice but also with a number of similar sounding voices serving as foils or distractor voices. The series of voices is referred to as a "voice lineup" or "voice parade." As in an eyewitness lineup, the witness has to decide if he/she recognizes the voice of the criminal from the voices in the lineup. Different procedures for this type of identification exist among countries and also within the same country. The Committee for Standards in Earwitness Lineups was formed by the International Association of Forensic Phonetics (IAFP), to set up guidelines for earwitness identification procedures (Hollien et al., 1995). Since then, many workers in the field have published recommendations or research on voice lineups (Broeders, 1996; Broeders and Rietveld, 1995; Hollien, 1996, 1997; Hollien et al., 1995; Künzel, 1994; Nolan and Grabe, 1996; Yarmey, 1995).

This study will investigate the earwitness aspect of voice lineups. The next section is concerned with what is known in this area. First, however, a summary is given of the

general factors discussed earlier, that also applied to voice lineups: 1) The quality and length of the speech samples used can influence accuracy significantly, where high quality and longer recordings improve identification performance. 2) Other research suggests that disguise is a powerful tool for the speaker that does not want to be recognized, as it severely degrades accuracy. 3) To a lesser degree also temporary speaker distortions like a cold may affect the correct identification score. Where the speaker is concerned: 4) certain voices are easier to recognize than others. A unique voice will be remembered for a longer time than a voice lacking that characteristic. 5) The fact that voices change slightly over time, seems to affect only those cases with a latency above six years. Fortunately, this means that the “non-contemporary” problem only applies to a minority of investigations. 6) Where it concerns the listener, it was found that there exists a wide range in natural speaker identification skills. Some almost seem to have a special developed skill for correctly identifying speakers while others show a performance at the other end of the spectrum. 7) It turns out that training in (forensic) phonetics has a positive influence. 8) Research on gender differences is inconclusive, but the age of the earwitness does have an effect. 9) Studies have shown that one should be careful when the witness is a child or an older adult. 10) The same care should be exhibited when the listener does not speak the language or dialect of the speaker. Please note, that familiarity with the speaker does apply neither to the regular voice lineups (Broeders, 1996; Wagenaar, 1988) nor to the focus of this study. In the case of a parade, it is assumed, that the witness was not previously acquainted with the offender’s voice. In the case of a familiar speaker, the identification of the offender has already taken place, either at the

time the crime was committed or directly after (Broeders, 1996). The remaining factors that also apply to voice lineups and earwitnesses are the following:

Latency between first confrontation and identification. Compared to eyewitness research, where even a few months-delay does not significantly affect correct identification, latencies related to earwitness identification are much more of an important factor. Results very much depend on the setup of the research (open/closed set, length signal, number of distractors, quality speech sample, etc.), but the overall factor of “time” is more detrimental in the area of earwitnessing than in eyewitnessing. For example, McGehee found a considerable decay after the first week. Her correct identification scores were as follows: 1-wk 81 %, 2-wk 69 %, 4-wk 57 %, and 20-wk 13%. In her 1944 study where recorded voices were used instead of live talkers, a significant drop was shown to occur within only two weeks, namely from 85% after two days to 48% after two weeks. If these scores were represented in a graph, the curve would very much resemble the Ebbinghaus forgetting curve, reported in general memory research (Ebbinghaus, 1885; Wixted and Ebbesen, 1991). It shows that the greatest amount of forgetting occurs in the first period after learning, with progressively less and less of a loss over time. McGehee’s scores might have been affected, however, by the modest quality of recording equipment of that time (1944). A surprising improvement was found by Hollien et al. (1983) and Saslove and Yarmey (1980). Hollien et al. tested 11 listeners (using seven foils) on day-1, week-1 and week-2. The trend, while not significant, appeared to be in the “wrong” direction with accuracy increasing from 36%



on day-1 to 50% after two weeks. Brown's explanation (1979) for this phenomenon is that identifications associated with long-term time intervals may be easier to organize than those associated with short intervals. Even though some researchers found an increase in accuracy over time, a majority of researchers report that identification accuracy declines as a function of time (Clifford et al., 1981; Clifford and Denot, 1982; McGehee, 1937, 1944). Here again, latency is a crucial factor and a decreased identification accuracy can usually be expected after two weeks (Clifford and Denot, 1982; McGehee, 1937). In general, it is therefore crucial to have witnesses perform identifications as soon as possible after the time of the crime.

Similarity of foils to target voice. Research on the construction of voice lineups has shown that if they include foils exhibiting voices very similar to the criminal's, identification is quite difficult (Broeders, 1996; Hollien et al., 1983; Rothman, 1977; Stuntz, 1963), increasing the number of false alarms (Handkins and Cross, 1991). Rothman (1977) used sound-alikes (e.g., father, son) and found that accuracy dropped significantly in a same-different task (from 94% with the standard foils to 58% with sound-alikes). However, this does not necessarily need to be considered as a "negative" phenomenon. In the case of a lineup, it should actually contain foils that are similar to the target: the use of speakers of a type quite different from the suspect is not an acceptable procedure (Broeders, 1996; Hollien, 1997).

The number of voices. The number of voices in the array is also a factor of consideration: Pollack et al. (1954), Clarke and Becker (1969) and Carterette and

Barnebey (1975) found either the correct identification rate or the false alarm rate to be negatively affected by an increase in the number of foils. In "Criteria for earwitness lineups" (Hollien et al., 1995) it is therefore suggested to limit the number of foils; utilization of a parade with too many speakers can tax witnesses' memory.

Earwitness' assumptions. Problems resulting from listeners's assumption that the criminal or his voice must be amongst the persons/voices in a lineup has been recognized by several researchers (Bull and Clifford, 1984; Hollien et al., 1983; Malpass and Devine, 1983; Warnick and Sanders, 1980). Hollien et al. found that in their study, innocent talkers were selected as the criminal a majority of the time for all trials. Few listeners took the option of indicating that the criminal was not in the group. The same has been found in earwitness identification (Malpass and Devine, 1981). Therefore, it is crucial that the agent carrying out the voice lineup informs the witness that the lineup may not contain the alleged criminal.

Earwitness' confidence. General memory research (Murdock 1974) seems to support a positive relationship between confidence and correctness: a person is more likely to be correct when he or she is certain of being correct. The same is true in the area of speaker identification (Clifford, 1980; Künzel, 1990; Rose and Duncan, 1995; Saslove and Yarmey, 1980; Thompson, 1985a). However, Hollien et al. (1983) with 58 listeners found the level of confidence in the identification to be lower for correct responses than for the incorrect. Also, Thompson (1985a) states that although there is an overall positive

correlation, subjects making incorrect identifications can be very confident about their choice. In his study, 47.8% of the subjects making incorrect judgements gave their choice a confidence rating of three (=highest confidence). Therefore, no conclusions can be drawn from a highly confident witness, even though, in general, a positive correlation exists.

### Summary

The factors reviewed above received considerable attention in the past 40 years. Certain issues, however, have not been investigated at all or have only been discussed superficially. Three of those issues are identification accuracy as a function of the listener's 1) memory, 2) auditory skills, and 3) musicality.

### Aural Perceptual Relationships Which Have Not Been Studied

As discussed earlier, many researchers in this field have observed that their listener subjects show a wide range of speaker identification skills (Bartholomeus, 1973; Bull et al., 1983; Coleman, 1973; Hollien and Thompson, 1990; Hollien et al., 1995; Iles, 1972; Künzel, 1990, 1994; Stevens et al., 1968; Thompson 1985b). One subject will score very poorly whereas another will seem especially skilled in the task. Even though many investigators are concerned with this relationship (Hollien et al., 1995; Stevens et al., 1968), research is lacking about speaker identification accuracy as a function of the specific and innate characteristics of an earwitness. So far, much of the reported research has focused on factors outside the auditor (e.g., the speaker, the speech material, construction of the lineup, etc.). When features concerning the witness were included, they were rather simplistic, for example, age and gender. Indeed, no specific memory or

hearing tests have been applied. Because voice lineups are carried out on a daily basis throughout the world, without these safeguards, investigating both relationships seems crucial.

### Memory of the Earwitness

Research on earwitness identification and memory. To date, no research has been carried out on speaker identification accuracy as a function of memory skills of the listener. Indirect features, however, have been studied, e.g. the effects of time-delay on memory, familiarity of the voice, uniqueness of the voice, etc. They have been discussed in detail in the former section. Another article discussing the relationship between earwitness identification and memory was from Brown (1979). In his paper, he explains the parts of the memory involved in experimental speaker identification tasks. Short-term memory tasks involve the known voice sample being presented after a delay of at most a few minutes from the presentation of the unknown voice sample. Same-different tasks are examples of this type. In long-term memory tasks, the target voice sample is presented after a longer delay. In this case, the known voice was already stored in long-term memory either through rehearsal earlier in the experimental session or because the pattern was already in long-term memory before the experimental session. Tests with familiarized voices (Williams, 1964; Stevens et al., 1968) and familiar voices (Pollack et al., 1954; Abberton, 1974) can therefore be called long-term memory speaker identification tests. In the case of familiarized voices, the listener is tested on memory of voices he/she heard and learned earlier in the experimental session. At that time, the

voices were stored in long-term memory. In the case of familiar voices, the patterns are already in long-term memory before the experiment. Brown also states that real-world earwitness cases always involve long-term memory. This means that to investigate earwitness identification, the experimenter should primarily focus on long-term memory, because it is the most important subsystem of memory used.

### Research on Memory in General

One of the more popular models in this area has been developed by Atkinson and Shiffrin (1965, 1968, 1971); it is sometimes called the “modal model” (Searleman and Herrmann, 1994). The general structure of memory is assumed to consist of a sensory memory, a short-term memory and a long-term memory. Sensory memory appears to hold the information from our senses (image, smell, voice, etc.) but operates for very short periods, that is, perhaps for less than a second. The image in sensory memory fades quickly and most of the information never proceeds beyond the sensory register.

Information that is selected for further processing is then transferred to the short-term memory (STM) which holds the contents of one’s attention. The material in STM will decay within 15-30 sec. (Loftus, 1980; Reed, 1973) unless one consciously attends to information. An example of this process would be repeating a telephone number several times until it can be written down. After that, it is either forgotten or it is transferred to long-term memory (LTM). This time span may seem limited, but it is probably more efficient this way, as it could create an overload. Typically, STM cannot hold on to seven items, plus or minus two, at one time (Miller, 1956). This result is known as a person’s

“memory span” defined as: the maximum number of items correctly recalled in order. Long-term memory is more or less thought of as a permanent storehouse of data. It contains all the events of a life-time. Further, there appears to be no risk of overloading long-term memory as it is considered limitless (Searleman and Herrmann, 1994). The preferred code of memory is semantic, but visual or acoustic coding also may occur in LTM (Gernsbacher, 1985; Sachs, 1974; Searleman and Herrmann, 1994). Interference effects are assumed to be the major cause of forgetting in LTM: other information or events can disturb or interfere with retention. There are two major forms of interference. Retroactive interference refers to newer information acting backward in time to cause disruption. The proactive type refers to previously learned information acting forward in time to cause disruption. Aging and neurological factors, also, may induce trace decay in LTM (Squire, 1987).

The question may be asked as to how memory changes as people get older. Researchers have found that sensory memory is hardly affected ( Craik, 1977; Crowder, 1980; Kausler, 1991). However, impaired cognitive processing ability does affect both STM and LTM (Burke and Light, 1981; Craik, 1977, 1987; Guttentag, 1985; Welford, 1958). For example, the “digit span forward” task does not decrease for individuals of older ages, but this task backward does (Bromley, 1958; Mueller et al., 1979). Also, older subjects are particularly poor at tasks requiring free recall of information from LTM as apposed to tasks requiring just recognition (Botwinick and Storandt, 1974; Craik, 1977; Craik and McDowd, 1987). For LTM this means that free recall will be impaired as a function of aging, but recognition may not.

Considering earwitness identification, it can be assumed from the above that it is mainly LTM that is involved -- especially when it is compared to same-different discrimination tasks (Brown, 1979; Hecker, 1971) wherein STM is the subsystem mostly involved. Of course, STM is also employed when identifying a speaker, but the LTM has a more dominant role.

Many researchers have tried to distinguish different types of LTM (Searleman and Hermann, 1994). Some believe in a semantic/episodic memory distinction. The semantic memory is defined as the database for general or generic knowledge about the world, like symbols, rules and facts. The episodic memory element is involved with specific events and experiences relative to a person's life; they are all autobiographical events. Others have considered the declarative/nondeclarative memory distinction more useful. Declarative memory stores data that can be acquired in a single trial and that are directly accessible to conscious recollection (e.g., learning new words in Spanish). Nondeclarative memory deals with learning that is obtained incrementally and that is inaccessible to conscious recollection (e.g., learning how to swim).

From the above, it can be concluded that episodic (and declarative) memory is involved when remembering a heard voice. Also, if an earwitness does not remember the voice, two things may have occurred. First, memories of the voice may have never reached the listener's LTM and therefore cannot be found. The second possibility is, that the features of his/her voice were stored in LTM, but just cannot be retrieved. The fact that memory retrieval deteriorates with age, explains why young individuals perform better when identifying speakers than people of older age.

As it turns out, the processing of episodic memory is governed by different principles. Time affects memory in such a way, that recent experiences are recalled better than old ones. The retention function was found to decrease monotonically (Rubin et al., 1986) and that is true for all adult age groups: most recent personal memories will be recalled. The greater an event's latency, the less likely it will be remembered. This decrease over time explains McGehee's decaying curve for speaker identification accuracy. On the other hand, reminiscence accounts for the results with older subjects, who show an increased tendency to recall events from their lives that occurred when they were 10 to 30 years old. Searleman and Herrmann (1994) suggest that these memories are often thought about by people and are thus preferentially sampled by older adults. The crucial question in regard to witness identification of course is whether autobiographical memories can be trusted. Unfortunately this is not always the case: while most of the data stored in autobiographical memory is usually accurate, our personal memories are susceptible to systematic distortions, especially in terms of their fine details (Barclay, 1986; Brewer, 1986; Conway, 1990; Linton, 1986; Neisser, 1981). Memory is reconstructive in nature (Bartlett, 1932), a process which may lead to faulty remembrance of personal memories. Gaps are filled in with details one believes must have happened on the basis of plausible inferences (using general scripts or schemas). Neisser (1981) called it "repisodic memory" which is blending together details from many similar episodes. Even worse, if misleading information about an event is presented, people often have difficulty remembering the original event. This is called the "misinformation effect" and there is extensive evidence: (e.g., Belli, 1989; Ceci et al., 1987a, 1987b; Lindsay, 1990;



Loftus, 1975, 1977, 1979a, 1992; McCloskey and Zaragoza, 1985; Tversky and Tuchin, 1989).

Juries usually place an inordinate amount of trust in eyewitness accounts (Searleman and Herrmann, 1994). It is not known to which degree these data apply to earwitness identification, but the information about eyewitness testimony shows that memory is not a stable entity. Loftus (1986) even states that it is estimated that there are about 8500 wrongful convictions each year in the United States and that perhaps as many as half of them were the result of faulty eyewitness testimony. Although both types of identification are well accepted by the courts in the United States, great care should be exhibited in cases of witness testimony (DeJong, 1996; Loftus, 1986; Hollien et al., 1983).

The earliest battery for testing memory was published by Wells and Martin (1923). It included some 26 items, and the range of cognitive and memory tasks included suggests that it was as much a test of mental efficiency as of memory. Similarly, the Babcock Test of Mental Efficiency (Babcock and Levy, 1940) included learning and memory tasks in an extensive battery of tests of mental efficiency. In 1945, Wechsler published his Wechsler Memory Scale which later was revised in 1987. It was designed to investigate memory or memory loss in populations between the ages of 16 to 69. The functions assessed include memory for verbal and figural (visual) stimuli, meaningful and abstract material, and delayed as well as immediate recall. The 12 subtests are grouped under five separate memory scores: verbal memory, visual memory, general memory,

attention/concentration and delayed recall. The reviews of the WMS (and WMS-R) were positive from the beginning and the test still enjoys great popularity (Erickson and Scott, 1977; Gass and Apple, 1997; Hamsher, 1990; Holden, 1984; Kogan, 1949; Kreiner et al, 1997; Mayes, 1995; Mensh, 1953, Thompson and LoBello, 1994). Moreover, it has been shown to exhibit validity and reliability (Prigatano, 1978; Russell, 1975, 1981). Indeed, the Revised WMS is considered to be the most stable and valid memory test battery available (Searleman and Herrmann, 1994) and it is considered the most appropriate test for the major domains of verbal and visual memory (Holden, 1984). One disadvantage may be that the test is clinical and mainly indicates memory impairment; that is, it may be less sensitive to differences in persons with normal memory. However, considering both the advantages and disadvantages, the WMS was still considered the most appropriate test in regard to this research.

### Auditory Skills of the Earwitness

Research on earwitness identification and auditory skills. Surprisingly enough, very little literature can be found on how auditory competency relates to speaker identification even though such skills are crucial in the case of earwitness identification. In many studies, hearing tests are performed to ensure that the listeners can carry out certain aural experiments, but the auditory assessment is usually not the focus. However, currently, more papers have appeared pointing out explicitly the need for an adequate hearing ability on the side of the witness (Künzel, 1994; Hollien, 1995). Künzel's suggests a thorough examination by an audiologist when symptoms of a hearing loss can be observed. One of

the reasons is that certain speech characteristics important for identification may be in the frequency domain where the witness has a hearing loss (e.g., a lisp). In his paper about guidelines for earwitness lineups, Hollien (1996) states that “it is important that a witness can be shown to be competent to carry out the task. To do so, they should be able to demonstrate that they: a) Heard and attended to the perpetrator’s voice, b) Have adequate hearing for identification purposes, and c) Show the competency to respond appropriately to the task.”(Hollien, 1996, pg. 18). Both papers indicate that there exists a need for more research on the relationship between auditory function and earwitness identification.

Research on auditory skills in general. A number of tests have been developed to assess a person’s auditory sensitivity. For example, one can determine auditory sensitivity to frequency by measuring the intensity required for a listener to detect the presence of a sinusoid at each of many frequencies. This test is called the puretone air conduction test. In this study, it will be used to ensure that each listener included in the study does exhibit normal hearing.

Two tests can be used in order to test sensitivity to frequency. One is known as frequency discrimination. The second test is called frequency selectivity or resolution. The first one measures the least perceptible difference in frequency that the subject can detect. To do so, two tones are presented one after the other and the listener has to indicate whether there is a difference between the them. The least perceptible difference can be very small, perhaps as small as 0.2% or 0.3% of the stimulus frequency (Pickles,

1988; Sek and Moore, 1995). It tends to be smallest for middle frequencies (around 500 Hz) and larger for very high and very low frequencies (Sek and Moore, 1995). Frequency discrimination results, however, are greatly affected by the method used (Complex vs. Sinusoidal, order of tone presentation), but the pattern stays the same. This type of sensitivity for frequency is further discussed and tested in the music section (Seashore test: Pitch discrimination).

The second test, frequency selectivity (also called frequency resolution), measures the extent to which the subject is able to filter one stimulus out from others on the basis of frequency. The subject has to detect one frequency component of a complex stimulus in the presence of other frequency components, all presented simultaneously. It has been demonstrated that this type of frequency selectivity plays an important role in many aspects of auditory perception (Moore, 1997) and, in particular the perception of speech (deBoer and Bouwmeester, 1974; Bonding, 1979; Evans, 1978; Dreschler and Plomp, 1980; Horst, 1987; Ritsma et al., 1980; Tyler, 1979). Specifically, frequency selectivity tends to correlate with speech perception in noise (Festen and Plomp, 1983; Horst, 1987; Tyler et al. 1982), a task that has been shown to require many of those analyzing abilities which are thought to overlap with the skills required for speaker identification (Köster et al., 1997). For example, subjects that are able to perceive speech very well even under those circumstances where competing noise is present, also often exhibit a high frequency selectivity. Second, frequency selectivity is important for identifying small details in the speech signal in order to understand it correctly. For example, it is crucial for accurate phoneme recognition (Dreschler and Plomp, 1980; Patterson et al., 1982; Festen and

Plomp, 1983; Stelmachowicz et al., 1985). It (frequency selectivity) also may play a role in judging voice quality (Rosen and Fourcin, 1986). For example, in the case of a “creaky voice,” vocal fold vibrations are of low frequency and although regular can give the percept of a “rough” or even diplophonic (two-pitched) quality. Since the skill to select a frequency is crucial for accurate phoneme recognition, as stated earlier, it also may be important for other details within the spectrum that people are able to identify: for example, vowel formant frequency level, ratios, and transitions (Evans, 1978; Ices, 1972; Meltzer and Lehiste, 1972; Scharf, 1978; Stevens et al., 1968). In short, it is assumed that, if frequency selectivity is an indicator of a listener’s sensitivity for details in the acoustic signal (e.g., speech), that it also may be an indicator for identifying small acoustic differences between the voices of different speakers. For example, the skill may be important in detecting differences in the spectra created by differences in the vocal tract (Joos, 1948; Peterson and Barney, 1952) or in the glottal characteristics (Monsen and Engebretson, 1977). How those characteristics can effect the speech spectrum (or quality of the voice) has been nicely described by Nolan (1983) and in less detail in Laver (1980).

Another auditory assessment associated with detecting small changes in the acoustic signal is temporal resolution or gap detection. It refers to the ability to detect changes in stimuli over time. For example, it refers to the ability to detect a brief gap between two stimuli or to detect that a sound is modulated in some way (Moore, 1997). As pointed out by Viemeister and Plack (1993), it is also important to distinguish the

rapid pressure variations in a sound from the slower overall changes in the amplitude of those fluctuations. In other words, it is an indication of a person's resolution of changes in the spectral envelope of a signal: this ability should be crucial to speaker identification, because differences in voices result in different spectral shapes (Laver, 1980; Nolan, 1983). Temporal resolution also has been frequently shown to be related to speech perception (Glasberg and Moore, 1989; Irwin and McAuley, 1987; Tyler et al., 1982).

Loudness perception is measured in terms of the least perceptible difference in the strength or loudness of sounds that the subject can detect. Humans are able to detect relatively small changes in sound level (0.3-2dB) for a wide range of levels and for many types of stimuli (Moore, 1997). For pure tones, discrimination performance improves with increasing sound level up to about 100 dB SPL (Riesz, 1928; Jesteadt et al, 1977; Viemeister and Bacon, 1988). Even though vocal intensity has not been investigated very much, it nevertheless is assumed that intensity level and variability are both used in speaker identification (Hollien and Köster, 1996). The parameter is further discussed in the section for music (Seashore test: Loudness discrimination).

Several different tests have been developed for assessing a human's ability to analyze speech, which activity must be considered a more complicated type of stimulus. Two major types of tests exist in this regard, the first of these being a threshold measure for speech understanding (Katz, 1994). In the first case double syllable words with a spondaic stress pattern are the most commonly used material. The speech recognition

threshold (SRT) is the intensity level at which the listener can repeat (or otherwise indicate that the word was recognized) 50% of the material presented.

The second type of speech test is a supra threshold measure that uses monosyllabic words (with or without background noise) to determine the listener's ability to understand speech. The result in this instance is frequently referred to as the speech discrimination or word recognition (WR) (Katz, 1994). Initially, tests of this type were used to evaluate the intelligibility of communication systems (House et al, 1963; Nixon et al, 1982; Hagness, 1970; Williams et al, 1965). However, their use later on was also extended to research of a more clinical nature (Elkins, 1971; Nabelek and Mason, 1981; Stark and Hagness, 1972; Swain, 1972; Williams, 1982). They were found to show good reliability (House et al, 1963; Stark and Hagness, 1972; Williams et al, 1965). The results of this type of test constitute a good indication of the listener's central auditory processing skills (CAP). In turn, CAP is concerned with the efficiency of using the auditory system to carry out complex processes. Good examples here would be discriminating speech, background noise suppression, locating the source of sound and integrating auditory information with that from other modalities (Katz, 1994). When a stimulus is presented, a listener first has to detect it and subsequently process it. A CAP deficit is present when the individual is not able to make full use of the heard signal. This problem has been associated with learning disabilities, especially reading difficulties (Monroe, 1932; Orton, 1964; Sawyer, 1981). Also, problems with phonics and reading comprehension, spelling, articulation problems, and poor communicative skills are characteristics of individuals

with CAP deficits (Bannatyne and Wichiarajote, 1969; Boder, 1973; Fried et al., 1981; Kavale, 1981; Mange, 1960; Stovall et al., 1977).

It is generally assumed that many of the analysis skills required for identifying speech also are required for identifying speakers (Hollien and Köster, 1996). Actually, very little research has been carried out which addresses this relationship. About the only one was the report by Köster et al. (1997), which showed a positive correlation between speech perception and speaker identification.

### Musicality of the Earwitness

Research on earwitness identification and musicality. It seems reasonable to assume that earwitnesses use prosody as identification cues (Hollien and Köster, 1996); accordingly, it also would appear that musical skills, training, or talent could be important to the process. The two published studies that have been carried out investigating this relationship were very limited in size (i.e. with a sample size less than 7), but they do show a correlation between musicality and identification of speakers. For example, McGehee 1944 used three different types of individuals to analyze her speakers' recorded voices on the basis of pitch, rate of speaking and agreeableness. The groups consisted of individuals who were 1) trained in speech, 2) music and 3) not trained in either field. From five voices they had to indicate which voice was most unlike the others. When judging both pitch and rate, the listeners from the Music Group outperformed the ones from the other two groups. The same was true for the second part of the study; here, the listeners were asked to judge the speakers' age, height, weight, and personality



characteristics (introversion-extroversion and ascendance-submission) on the basis of voices heard. She concluded that on the whole, the people trained in music seemed to be better judges than either those trained in speech or without special training. Also Köster et al. (1997) found that the musicians in their study scored highest or very high in the speaker identification task. However, no statistical significance could be reached due to the very small sample size. On the other hand, Shirt's investigation (personal communication, 1996) did not show a relationship between musicality and speaker identification accuracy. It seems that more research is needed to obtain a clearer definition of this relationship.

Research on musicality in general. Musicality or musical aptitude implies the potential, usually innate, for developing musical skills. Therefore, it includes those factors which are not influenced by training (Hodges, 1980). The first standardized test of musical aptitude -- and the best known -- is the Seashore Measures of Musical Talent (Seashore, 1919). However, many tests have been developed afterwards. Some examples are the Kwalwasser-Dykema Music Test (1930), the Drake Musical Aptitude Test (1933), the Wing Standardised Tests of Musical Intelligence (1948), the Tilson-Gretsch Musical Aptitude Test (1941) and the Gordon Musical Aptitude Profile (1965). Almost all tests have been upgraded to recent versions.

What do these tests measure, or, in other words, which factors are known to indicate musicality? As it turns out, almost all musical tests include measures of pitch,

tonal memory and rhythm and most of them assess, in addition, timbre/quality and loudness. Can these features be related to speaker identification at all? Indeed, many musical parameters, although not thoroughly investigated as yet, could be logically linked to this ability.

Pitch discrimination, for example, may be related to speaker identification for several reasons. First, due to a certain degree of sensitivity people are able to perceive a speaker's pitch and pitch variability. Research in speaker identification has shown that those features can be identified and used as a cue for identification (Compton, 1963; Iles, 1972; LaRiviera, 1971). They are able to identify the speaker as a man (100-130Hz) or as a woman (190-220Hz) or even as a child may be (300 Hz at 10 years of age) (Fant, 1956; Hirano, 1981). Within the spectrum of a particular gender, they can also specify whether a person speaks with an unusually high pitch or that someone's pitch is higher than someone else his/her pitch. The ability to identify a speaker's intonation is another factor related to this parameter: defining someone's intonation would not be possible without the ability to define the speaking fundamental frequency of the talker.

Tonal memory tests how well listeners can remember a sequence of tones. This of course, should correlate highly with the talent to perceive a speaker's intonation: here also, a listener has to identify and remember shifts in frequency. Indeed, research has shown that intonation and prosody are features that differ among speakers (Darwin and Bethell-Fox, 1977) and that can be identified (Hollien and Köster, 1996) and remembered by listeners (Church and Schacter, 1994).

In rhythm tests, subjects are usually required to compare two rhythmic patterns and indicate whether they are the same or not. As with tonal memory and pitch discrimination, this parameter most probably is related to the perception of intonation and prosody.

The purpose of a timbre/quality test is to measure a person's ability to discriminate between complex sounds which differ only in harmonic structure. Identifying the quality of tones seems to be a task that is similar to the perception of the quality of a voice. Since voice quality has been shown to be one of the most robust parameters in automatic speaker identification systems, an individual's sensitivity to tone quality also may be related to the identification of voices. The voice parameter, that is, the long term spectrum, was found to show high accuracy levels, and is also resistant to the effects of speaker stress and to limited passband conditions (Bricker et al., 1971; Clarke and Becker, 1969; Doddington, 1970; Furui, 1978; Kosiel, 1973; Majewski and Hollien, 1974; Zalewski et al., 1975).

Even though the relationship between vocal intensity (or loudness) and speaker identification has not been investigated in detail, it is assumed that intensity level and variability are both useful as cues to speaker identification (Hollien and Köster, 1996). Speakers do differ in their level of speaking intensity, and this parameter has been successfully used in automatic speaker recognition systems (Doddington, 1971; Lumnis, 1973; Rosenberg and Sambur, 1975). Further, in a related area, Scherer (1974) found that lay listeners primarily use pitch and loudness to judge vocal quality and Glasberg and Moore (1989) found that intensity discrimination is one of the factors that can predict a person's speech understanding ability.

Several other relationships concerning the musical aptitude of an individual have been investigated. Do individuals who are talented in this area also show superior aural acuity? In a psychoacoustic study, Webster et al. (1950a,b) found that non-musicians had a poorer frequency selectivity than musicians. Sergeant (1973) gave musicians and nonmusicians a pitch discrimination task with both pure and complex tones. The musicians were superior to the nonmusicians in both tasks. The fact that musical ability and aural acuity are related was also suggested by Farnsworth (1941) Tomatis (1953), and Nass (1990).

### Objectives of This Research

The focus of this research is on earwitness identification. The main question asked is, why are certain people very good at identifying speakers while others perform poorly at this task. In other words, what are the features that distinguish the individuals that are good at identification from the ones that are not. In order to study the cited relationships, the method of extremes will be used. That is, a group of subjects that show good earwitness identification skills will be contrasted to a group that exhibits low skills. The two groups will be formed by administration of a procedure where the speakers are unfamiliar to the auditors and where a time-lag exists between the first speaker-auditor confrontation and the identification of the voice by the subject who is now serving as an earwitness. That is, the construction used in this study reflects the setup employed in real cases. Again, the goal of this investigation is to find those parameters that are important for speaker identification. Accordingly, it is hypothesized that:

- 1) the quality of memory skills affects a person's ability to identify speakers; that is, poor memory skills operate to decrease, and good skills increase, an individual's identification accuracy,
- 2) the level of auditory skills affects a person's ability to identify speakers. Again poor audition skills are expected to decrease, and good auditory performance to increase, an individual's speaker identification accuracy, and

- 3) musical skills also operate to affect a person's ability to identify speakers; that is, poor musical skills will decrease and good skills increase an individual's accuracy.

## CHAPTER 2

### METHOD

As stated the objectives of this study are to identify individuals who are either particularly good or poor at earwitness identification. Next, it is determined if certain memory and auditory skill factors correlate with identification performance and to what extent they do so. Even though very little research has been carried out that assesses these (earwitness) attributes, this type of identification is performed on a regular basis all over the world. Since witness characteristics are crucial to the procedure, good information about them appears needed. To ensure that the results of this study would indeed apply to realistic earwitness identification, it was considered extremely important that the identification test employed in this study would resemble real-world cases as closely as possible. To meet the cited goals, certain procedures were established. First, listeners were selected to be unfamiliar with the speakers; second, the first confrontation with the voice was structured in such a way that the listener “unintentionally” overheard it. That is, the listener was not instructed to pay specific attention to the voice itself and was not informed that she would be required to identify that particular voice later on. Finally, a time-lag of 2 weeks (plus/minus 1 day) existed between the first confrontation and the identification task. Since this latency was, perhaps, minimal, the approach was made more difficult by requiring multiple speakers to be identified.

## Subjects

### General Subject Selection

The initial subject pool consisted of 112 young female college students enrolled at the University of Florida. The subjects were between the ages of 18 and 35 years with the mean age of 21. They all reported American English as their first language and exhibited a general American English dialect. Female subjects were chosen, because attempting to study both sexes would be too work intensive and because most of the earwitnesses are female in real-world cases. Moreover, gender differences already have been studied (McGehee, 1937, 1944; Thompson, 1985a). The subjects were recruited from both graduate and undergraduate classes in Linguistics, Audiology, and Speech Pathology. None of them received payment for participation, but about 65% received class credit. The cited age range was selected because it permitted an initial focus on young adults as well as the avoidance of those variables (i.e., puberty, aging, etc.) which could confound the results obtained in the study. Subjects could only participate if they 1) agreed to sign the “Informed Consent Form”, 2) were in good health (medical questionnaire, see Appendix A) 3) exhibited normal hearing (medical questionnaire and puretone air-conduction threshold test), and 4) had not been diagnosed as suffering from memory related illnesses (medical questionnaire). The medical questionnaire consisted of selections that were drawn from documents used by the University of Florida Speech and Hearing Clinic and the Department of Clinical and Health Psychology.

Since this research focused on the study of subject capability regarding auditory sensory channels, only subjects with normal hearing could be selected. Hence, they were



screened for normal hearing at the University of Florida acoustic phonetics lab. Both ears were tested individually in an IAC sound-treated room which meets OSHA requirements for headphone testing. The pure-tone stimuli were presented monaurally to the subject via a portable audiometer (MAICO MA 40) with TDH-50 headphones and testing was conducted following ANSI guidelines (1991). Any subject exhibiting a hearing loss was removed from the subject pool. Hearing loss was defined as any threshold poorer than 25 dB HL at the frequencies of 500, 1000, 2000, 4000, or 6000 Hz (ASHA, 1990).

Of the 112 subjects, six canceled before completion of the voice lineups. In addition, two subjects were excluded from the study due to a hearing loss discovered after they had started participation. In short, 104 women served in the basic subject cohort.

#### Selection of LOW-SPID and HIGH-SPID Groups

To select the listeners for the experimental groups, all 104 volunteers that satisfied the general requirements for participation were subjected to a speaker identification test. It consisted of two parts, 1) confrontation with the speakers and 2) identification of the speakers. On day-1, each subject was confronted with tape recordings of four speakers, two male and two female talkers, that were unfamiliar to them. They were seated in the sound treated IAC-booth located in Dauer Hall at the University of Florida. Beyerdynamic DT 211 headphones connected to the TEAC (X300) were used to play the speech samples at a level the listeners thought was comfortable. Since, they were instructed to listen to different speakers discussing various topics and asked to remember

the content of each monologue, a situation was created in which they listened to the speakers without their voices being the focus of the process. Indeed, they were told that they would be questioned later about it as part of the "verbal-memory test." Immediately after they had heard all four speakers, they were informed that the above instructions had been false, and the real task would be to identify the speakers' voices in two weeks time.

After a period of two weeks (plus/minus one day), all subjects returned to the lab for the speaker identification task. They were told that they had to select the four speakers they had heard two weeks before, from a series of voice lineups. Each speaker had to be identified in five different lineups. In eyewitness identification the witness is permitted almost an endless number of trials (30-40 trials per typical lineup) and the order of the trials is unstructured. However, in this study only five lineups were employed in order to control the number and keep the task organized and within reason. To ensure that the target speakers (i.e., Male-1, Female-1, Male-2, Female-2) were not confused with each other, the listeners were provided the details of the monologue uttered by each speaker. Confusion of the four speakers, however, did not appear to be a problem as all listeners indicated that they clearly understood which target speaker they had to identify. Listeners were also instructed to consider each lineup independently from the others even though the lineups consisted of the same speakers. It should be avoided that listeners would score all lineups correctly/incorrectly solely because of being consistent. As stated, the listener had to identify -- in turn -- two male and two female speaker: each target speaker had to be identified in his or her own set of five different voice lineups. Only one speaker had to be identified per lineup: the other four foils or distractor voices were unfamiliar to the

subject. Moreover, a different set of distractor voices was used for each target speaker. Since the procedure involved was a closed identification task, the target was always among the five speakers in the lineup. All subjects were informed about the fact that they always had to choose one and only one, since the target speaker was in every lineup (closed set). They were also told that the same set of foils were used for each target, but that each speaker had his/her own set of distractors. The highest score a witness could obtain for this test was 20; that is, a point for each of four speakers in five voice lineups. Thus, a score of 20 -- or 100% -- meant that the subject had identified all targets in the lineup correctly. The lowest score of course was zero. The highest score obtained in this study by any subject was 80% and the lowest score was 0%. A display of the results can be found in Fig. 1. The mean of the correct identification scores was 32% with a standard deviation of 17.5%. Overall, it can be seen that the scores are quite low with the mean being only about 10% above the 20% chance level. One explanation is that the task itself was perhaps too difficult. The listeners were only exposed to 30 sec. of speech when confronted with the target speakers and they had to remember four speakers, two men and two women. Confusion of the speakers may have been another reason. However, since the content of the monologue of each speaker was different for all four speakers and since the subjects associated each speaker with his/her story, listeners clearly indicated that they knew which speaker to identify. The fact that the scores were quite low is not necessarily a problem: the speaker identification task should be challenging enough in order to avoid ceiling effects. Second, for the method of extremes, a reasonably wide spread of the data is desired so that two extreme groups can be selected. The distribution here satisfied that requirement.

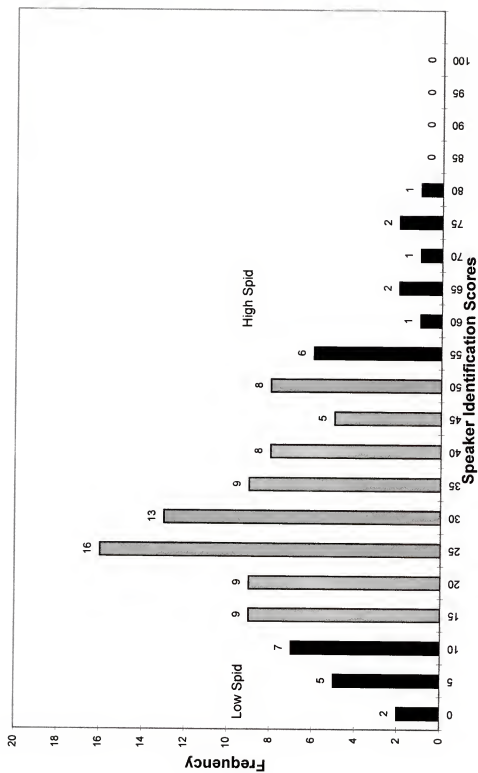


Fig.1 Histogram: SPID scores and their frequency

After all 104 volunteers had participated in the identification test, the experimental groups were selected. The High Group (or HIGH-SPID) consisted of those 13 who had completed the identification task (see above) and achieved high scores -- that is, values of 55% correct or better. The other cohort was made up of 14 subjects who scored 10% or lower on the identification task (Low Group or LOW-SPID). After some pilot research, it was judged that these thresholds would be the most appropriate. By employing these thresholds, two statistical requirements were satisfied. The first one requires a reasonably large distance between the two groups and the second condition demands a large enough sample size. Since the distance between the groups was 45% and the sample size 13, both requirements were met. The groups were very similar in age: the mean age for LOW-SPID was 21.7 yrs. and HIGH-SPID 20.1 yrs. Combined, they consisted of about one quarter (27 subjects) of the total group originally screened (104 subjects). This meant that 75% of the subjects was discarded. The mean SPID score for LOW-SPID was 7% and for HIGH-SPID was 63%.

Did the listeners in the LOW-SPID group perform worse than the HIGH-SPID group, because they were hesitant to alter their initial (and incorrect) decision in subsequent trials and vice versa? In other words, was locking in on the wrong speaker the reason for their low score and did the same effect explain why the HIGH-SPID group performed so well? An analysis of their selection patterns revealed that this was only partially the case: selecting the same (wrong) speaker all five times in the lineup set only occurred in 5% of the cases. Choosing the same incorrect speaker four times occurred in only 21%.

### The Speakers and Speech Samples

To construct the tapes for the speaker identification test discussed in the former paragraph, twenty talkers were drawn from the extensive database of speech samples at the Institute for Advanced Study of the Communication Processes (IASCP). Two tapes were constructed: the speaker-auditor confrontation tape (tape-1) and the speaker identification tape (tape-2). Tape-1 consisted of extemporaneous speech of four different speakers each 25-30 sec. long. This tape was played to the listeners on day-1. Two male and two female target speakers were used, so that both genders were represented in the identification task. Tape-2, that was played 14 days after the confrontation, consisted of 20 lineups, five per target speaker. Even though in eyewitness lineups it is permitted to observe a suspect as long as the witness desires, here a series of only five lineups was used in order to keep the task within reason. A series of five was used. The speech material used for this tape was read sentences (see Appendix B). As stated earlier, each speaker appeared exactly once in each of the five lineups. The lineup consisted of the target voice and four distractors. Since it is crucial that the lineups were "fair", that is, the foils resemble the target voice in their voice and speaking characteristics, (Broeders and Rietveld, 1995; Hollien, 1997), several mock witness identification tests were carried out. The panel of judges included in total three forensic phoneticians, one phonetician, three speech pathologists and two graduate linguistics students. They were asked to select the speaker that sounded different from the other speakers in his/her voice or speech features. However, making a choice was optional; they did not have to select one, if they thought that all speakers sounded very similar. The final results, based on seven auditors and 35

judgements per set, may be seen in table 1. Note that choosing a speaker was optional and that the value indicated as “total” is the percentage of times that the listeners actually chose one. The table shows that the judges only decided to select a talker about half of the time. The other 50% of the time, they considered the lineup to be fair. After five distractor voices were replaced, the lineups were finally considered to be well balanced. This means that with the final collection of speakers, all foils were selected in roughly equal numbers and the target speaker was not chosen a greater number of times. The highest score obtained for a target speaker was 23% for speaker H130 in Set III. Since this value was close to the chance level of 20%, it was considered small enough for the lineup to be fair. For both Set II and IV the target speaker was only selected 6% of all the times that judgements were made.

Table 1. Mock witness test results

| Set I       | %  | Set II      | %  | Set III     | %  | Set IV      | %  |
|-------------|----|-------------|----|-------------|----|-------------|----|
| <u>H134</u> | 9  | <u>A205</u> | 6  | <u>H130</u> | 23 | <u>F220</u> | 6  |
| A102        | 6  | A204        | 11 | H133        | 3  | F221        | 11 |
| A103        | 11 | A203        | 6  | H135        | 3  | F225        | 6  |
| H135        | 6  | A201        | 9  | H136        | 0  | F218        | 17 |
| M117        | 14 | F212        | 17 | H140        | 29 | F219        | 11 |
| TOTAL       | 46 |             | 49 |             | 58 |             | 51 |

Note: the underlined subject number indicates the target speaker; for clarity they are always placed first in this table.

### Assessments of the Selected Subjects

In order to find out why certain listeners are good at speaker identification and others are not, the method of extremes was used. It involves selecting two groups: one group that performs very well at the task of interest (in this case speaker identification) and one group with individuals that exhibits very low scores. Subsequently, the characteristics of both groups are compared in order to find out where the groups differ. The features that show differences that are statistically significant are considered to be important for the task of interest. In general, the method of extremes is employed when 1) the research is exploratory in nature, and 2) the testing is both cumbersome and time consuming. Since this research was, indeed, exploratory and since the individual assessments took up to three hours, it was judged that the method of extremes was appropriate.

If subjects satisfied the requirements for participation, being selected as a member of either the LOW-SPID or HIGH-SPID group, they returned to the lab for an approximately two-hour session. In Table 2. an overview is given of the tests administered.

### Memory Assessment

The first set of assessments consisted of measurements of memory. The latest version of the Wechsler test was employed to assess subjects' memory skills; it is the Revised Wechsler Memory Scale (1987). It was chosen primarily because it extensively tests many facets of this entity. Moreover, it has been shown to exhibit validity and



Table 2. Overview of tests

|                                      |
|--------------------------------------|
| Memory Skills Assessment:            |
| 1. Mental Control                    |
| 2. Logical Memory I and II           |
| 3. Verbal Paired Associates I and II |
| 4. Digit Span                        |
| 5. Auditory Priming                  |
| Psychoacoustic Assessment:           |
| 1. Speech Recognition in Noise Test  |
| 2. Frequency Selectivity             |
| 3. Temporal Resolution               |
| Musicality Assessment:               |
| 1. Pitch Discrimination              |
| 2. Intensity Discrimination          |
| 3. Rhythmic Discrimination           |
| 4. Timbre                            |
| 5. Tonal Memory                      |

reliability (Prigatano, 1978; Russell, 1975, 1981). Indeed, the Revised WMS is considered to be the most stable and valid memory test battery available (Searleman and Herrmann, 1994). This test focuses mainly on verbal memory (Prigatano, 1978). Not all WMS subtests were included in this study. That is, the visual memory module was removed, because only auditory memory obviously relates to speaker identification; so was the information and orientation module because the simplicity of that test would result in a ceiling effect for subjects with normal memory.

Finally, although it was considered useful to test episodic memory (because it relates especially to the type of memory involved in earwitness identification), it was not included because testing it has been shown to be very problematic. That is, although people tend to give consistent answers (Searleman and Herrmann, 1994), it is most difficult to check their validity.

Apart from the auditory priming test, the following were all included among the submodules of the Revised Wechsler Memory Scale (1987). Scoring for each of them was carried out in a parallel manner: a percentile equivalent was calculated from the raw score.

1. Mental control: This test evaluated for accuracy and smoothness of automatisms: the subject was required to recite a familiar series of numbers or letters. The first task was to count backward from 20 to 1, the second was to repeat the alphabet as quickly

as possible. The third task was to count by 3's beginning with 1 (i.e. 1, 4, 7, etc.). The maximum score for each test was 2 points. However, one point was subtracted for each error. The final score consisted of the summation of the results of the three tests.

2. Logical memory I and II: These subtests evaluated subjects' ability to immediately recall verbal ideas from two paragraphs. Is it possible, that verbal memory is correlated with speaker identification accuracy? In other words, would detailed memory of the verbal content of the conversation at the time of the crime, also enhance the listener's memory of the speaker's voice? So far, no studies have been reported investigating this relationship, but the fact that the person is able to repeat the message may reinforce memory of the voice specific characteristics associated with it.

The subjects had to listen to two short stories and try to remember each just the way each was said -- or as close to the actual words as could be remembered. A point was awarded for each item remembered correctly. After both stories had been read and subject's responses recorded, they were told to try not to forget them, as they would be questioned about them later. Then, after about 20 min. the subjects were asked to review these stories for the second time (Logical Memory II). A point was given for each item remembered correctly after which all points (acquired with both stories) were combined to obtain the final score.

3. Verbal paired associates I and II: This subtest involved verbal paired-associate learning ability. The results of both tests were not used on their own, but were solely used to calculate the verbal memory and delayed recall score.

The subject was read a group of eight word pairs, then was read the first word of each pair, and was asked to supply the second word from memory. One point was given for each correct association. Here also, the subjects were questioned about the same word pairs after about 20 min. (Verbal Paired Associates II). A point was given for each word pair remembered correctly. Twelve points could be obtained in total.

4. Digit span: This subtest assessed the limitations of the subject's short-term memory. Since there is an information flow from short-term to long-term memory (Atkinson and Shiffrin 1965, 1971), the abilities of the first type influences how it gets stored and what gets transferred to long-term memory (Schmajuk and DiCarlo, 1991). Hence, this process undoubtedly is related to remembering someone's voice. The two parts of the Digit Span subtests, Digits Forward and Digits Backward, were administered separately. On Digits Forward, the subject was read number sequences of increasing length and after each sequence, was asked to repeat it from memory. Length increased from three numbers in the first, to eight numbers in the last sequence. On Digits Backward, the subject was read similar number sequences and, after each sequence, was asked to repeat it backwards instead of forwards. There were two trials for each number length. The subject was awarded two points if both trials were passed, one point if only one trial was passed and zero points if she failed both trials.

After all values were obtained, yet three other scores could be calculated. For example, the overall score for Verbal Memory was constructed in the following way:

the score for Logical Memory I was doubled ( in accordance with WMS-R for obtaining weighted raw score values) and added to the one for Verbal Paired Associates I. Table C-5 of the Wechsler Scale provided the index score of Verbal Memory associated with the result of the earlier summation. It indicates a person's learning and immediate recall abilities of verbal material. As stated earlier, it may be important that a witness remembers the verbal content of the criminal's monologue, as it could reinforce memory of voice-specific cues.

For two other scores (i.e., Attention/Concentration, Delayed Recall), the percentile score had to be derived from the mean Z-score, because certain values were missing since not all Wechsler items were administered. For example, a Z-score has to be calculated if an experimenter desires to compare his/her means to the ones that are defined as the norms, and subsequently to express the results in percentile equivalent. A Z-score is the difference between the mean and the norm expressed in standard deviation (e.g., half the standard deviation, one third, etc.); thus a percentile equivalent can be calculated using standardized tables especially designed for that purpose.

The percentile value for Attention/Concentration was constructed by averaging the Z-scores of Mental Control and Digit Span. The amount of attention paid to the criminal scene was considered crucial to speaker identification. It was hypothesized that an individual with a high score for attention/concentration will remember more of the criminal's specifics than a witness with a low score. This assumption was based on the fact that increased attention has shown to increase memory performance (Alain and Woods, 1997; Kellog et al, 1996, Mulligan and Hartman, 1996; Norman, 1976).

The percentile score derived from the mean Z-scores of Logical Memory II and Verbal Paired Associates II is an indication of the person's Delayed Recall of verbal material. In both subtests, recall is tested after a delay of 30 min. As stated earlier, it may be important that a witness remembers the verbal content of a criminal's message, as it could reinforce memory of voice-specific cues. The tables for the Wechsler norms contain percentiles for different age groups. Values obtained from the scores produced by normal 20-24 year olds provided baseline materials to which the present subjects could be compared.

5. Auditory priming: This term refers to implicit memory and in this study priming for voices was tested. "Implicit" means that information was acquired unconsciously and without intention (Graf and Schacter, 1985; Schacter, 1987). It is the opposite of explicit memory that entails conscious recollection of previously studied information, as assessed by recall and recognition (Schacter and Church, 1992). It has been shown that explicit memory and priming are separately functioning phenomena (Ochsner et al. 1994; Roediger and McDermott, 1993; Squire, 1992). For example, amnesic patients, who often exhibit severely impaired explicit memory, can nevertheless show normal perceptual priming. Researchers have studied implicit memory by measuring the effects of priming (Jacoby, 1983; Kirsner et al., 1989; Masson and Macleod, 1992; Schacter, 1990; Squire, 1987), an effect which occurs when a certain task, e.g. stem completion, is facilitated because of implicit memory acquired from recently presented items. For example, Schacter (1984) showed that when subjects were given a stem-

completion task, they tended to respond with words that they had heard previously as part of a different test. He also proved that a test can be constructed in such a way that only implicit memory is assessed (Schacter et al, 1994). Do people implicitly -- or explicitly -- remember voices ? In several studies, it was found that subjects show voice-specific priming: they remember not only the word or sentences spoken, but also details concerning the speaker's voice or speech like intonation, fundamental frequency, or gender (Church and Schacter, 1994; Cole et al., 1974; Schacter and Church, 1992; Schacter et al., 1994).

The experimental set up was derived from Schacter et al. (1994) who studied voice-specific priming in elderly adults. The order was as follows: first the meaning test was given, after which a short distractor task followed. Next, subjects were administered the stemcompletion test. Subjects were told that they would be hearing a series of word beginnings and that their task was to complete each stem with the first word that came to mind. After a 20 min. delay, they were given the cued-recall test.

The subjects were seated in a sound-proof booth. Stimuli were played on a reel-to-reel (TEAC) coupled to Beyerdynamic headphones (DT211). Listeners heard the same list of 24 words spoken by six different speakers, three male and three female. Their task was to rate the number of meanings for each word on a four-point scale (1 = one meaning, 4 = four or more meanings). There were five seconds between items for subjects to make their ratings. Subjects then had to carry out the distractor task. In the present instance, they had to fill out a questionnaire with questions pertaining to their

hearing, memory and musical skills. After the distractor task had been completed (it usually took between 2 or 3 minutes), subjects were given an auditory stem-completion test in which the first syllable of the studied and nonstudied words was presented.

Subjects were then instructed to respond to the heard stimuli with the first word that came to mind. If they could not think of a completion for a stem, they were instructed to write down the stem only. There were seven seconds between the items for subjects to write down their answers. The studied and nonstudied words were matched for frequency, first letter, number of syllables, number of possible completions from the first syllable and length (Graf and Williams, 1987; Kucera and Francis, 1967). Half of the stems from the studied items were presented in the same voice as that heard during the study task, and half were presented in a different voice, one that always involved a change in the speaker's gender. To calculate the amount of voice-specific priming, the following procedures were used. First, the proportions of stems completed with studied items were calculated for both different speakers and for the same. However, a certain adjustment was made to correct for stem misperception: all proportions were computed by dividing the number of target completions that each subject provided by the number of syllables that they perceived correctly. Thus, for example, if a subject misperceived three items, her number of target completions would be divided by 9 instead of 12. Next, another adjustment was made to correct for items scored correctly by chance; the baseline proportion (the ratio for the non-studied items) was subtracted from both the studied-items ratios (same speaker, different speaker). The final step was to subtract the studied-items ratio for different speakers from the ratio for the same speakers. The



resulting value indicates exactly how much priming for voices occurred. For a more detailed description, see Schacter et al (1994) and Schacter and Church (1992).

### Psychoacoustic Assessment

The second series of auditory tests consisted of psychoacoustic measurements.

1. Speech recognition in noise test: It is hypothesized that many of the analysis skills required for identifying speech also are required for identifying speakers (Hollien and Köster, 1996). Actually, very little research has been carried out which addresses this relationship. About the only investigation that (indirectly) studied this connection was Köster et al. (1997). They showed a positive correlation between speech “sensitivity” and speaker identification; that is, subjects that scored high on this speech test also performed better when carrying out a speaker identification test. However, their test did not directly assess speech reception abilities in noise, but a person’s analytical sensitivity towards different elements of speech (e.g., pitch contour, voice onset, perception and repeating of nonsense syllables). It was decided to use a multiple choice speech reception test, since they are easy to administer and usually take a relatively short time. In such tests, the subjects are required to select and circle the word they just heard from a small group of, for example, six items. The fact that administration and scoring of such a test is very straightforward, may reduce the number of errors. These type of tests were found to show good reliability (House et al, 1963; Stark and Hagness, 1972; Williams et al, 1965). Since the listeners in this study all exhibited normal hearing, white background noise was added

to avoid floor and ceiling effects. The masker consisted of a broadband noise from 250 to 6000 Hz with equal energy per octave.

The speech stimuli used came from the Modified Rhyme Test (MRT), developed by House et al. (1963, 1965) and modified by Kreul et al. (1968). They consisted of 50 familiar American-English monosyllabic words. The word form was either consonant-vowel-consonant (CVC), consonant-vowel (CV), or vowel-consonant (VC). White noise was used at a signal to noise ratio of +6 dB: that is, the speech stimuli were presented at a 60 dB SPL level together with noise that was presented at a level of 54 dB SPL. Subjects were required to select, on their sheet, the word they thought they had heard from a group of six items. If they were not certain, they were instructed to make a guess. As stated earlier, the answer sheets were in multiple choice form with six words per ensemble. In all cases only a single initial or final consonant was varied: the remainder of the word was consistent with the other five items. Each word was uttered within the carrier phrase, "*Number - Circle the word again.*" Emphasis was placed upon typical, rather than exaggerated, articulation. The carrier phrase was chosen so that the test word would be preceded by a neutral vowel and followed by one. This was to reduce the coarticulation variation of the second formant that typically accompanies vowels displaced toward extremes of the vowel triangle (Ohman 1966a, 1966b). The speakers exhibited a General American dialect. Speech stimuli were played on a stereo cassette tape recorder (Sony, TC-RX 606ES) and subjects listened to the stimuli under TDH-50P headphones (Telephonics 29 6D 200-2) while seated in a sound-treated room (Tracor, Model RS 253C). The speech stimuli and the competing noise was separately attenuated, mixed,

amplified, and presented binaurally using a GSI-16 Audiometer. The results of this procedure took the form of the percentage of words subjects identified correctly.

2. Frequency selectivity: Frequency selectivity (also called frequency resolution) refers to the ability of the auditory system to resolve the sinusoidal components of a complex sound. It has been demonstrated that this type of frequency selectivity plays an important role in many aspects of auditory perception (Moore, 1997) and, in particular the perception of speech ( deBoer and Bouwmeester, 1974; Bonding, 1979; Evans, 1978; Dreschler and Plomp, 1980; Horst, 1987; Ritsma et al., 1980; Tyler, 1979). It is assumed that, if frequency selectivity is an indicator of a listener's sensitivity for details in the acoustic signal (e.g., speech), that it also may be an indicator of identifying small acoustic differences between the voices of different speakers.

Frequency selectivity was measured by a notched-noise masking paradigm (Glasberg and Moore, 1986; Patterson, 1976). This procedure is predicated by the assumption that a person's ability to separate the components of a complex sound depends mainly on the frequency-solving power of his/her basilar membrane. The narrower the shape of the auditory filter, the more sensitive the basilar membrane in resolving frequency. In the notched-noise procedure, the threshold of a sinusoidal signal is measured as the function of the width of a spectral notch in a noise masker. The shape of the frequency curve can be calculated measuring the thresholds at different notch widths. The notched noise paradigm appears to enjoy several advantages over other

techniques used to estimate frequency selectivity. First, it reduces the extent of off-frequency listening (i.e. the listener attending to more than one filter) which can lead to an apparent improvement in frequency resolution (Moore, 1982; O'Loughlin and Moore, 1981, 1986). Second, the procedure eliminates interactions between signal and masker, such as combination tones and beats, which contribute to inconsistencies in threshold patterns by supplying additional cues to the listener (Tyler, 1986). Finally, frequency selectivity can be estimated separately from processing efficiency, since the signal and the masker are separated, allowing for the separation of sensory from non-sensory factors. The thresholds obtained by using the notched-noise method are similar to those reported previously where the investigators used other paradigms (Dubno and Dirks, 1989, Moore and Glasberg, 1983b; Moore et al., 1990b; Shailer et al., 1990; Zhou 1995).

In the experimental set up of this research, the notch was positioned symmetrically around the signal frequency to assess the shape of the auditory filter for each listener. The thresholds were obtained for a 0.0 ms and 0.3 ms notch-width condition with the noise masker set at 50 dB SPL. The subjects were presented with two successive burst of noise, both interrupted by a small gap, where one of the two bursts contained a sinusoid. They were required to indicate the one containing the sinusoid, or in other words, the signal with the "chirp." The method of maximum likelihood estimation was used to obtain the frequency selectivity threshold. In other words, those values were chosen as estimates of the parameters that were most consistent with the sample data, or a  $\mu$  for which the likelihood value was largest. The number of trials was set at 30 and each subject's frequency selectivity threshold was measured three times for both notch-width

conditions. Before the official trials were started, however, a short training session was held until the subject clearly showed a good understanding of the procedure. This was considered necessary, since, during practice trials, it was found that listeners clearly needed familiarization with the unnatural and rough sounding stimuli. To obtain an impression of the shape of the listener's auditory filter, the averaged threshold of the 0.3 ms condition was subtracted from the one for the 0.0 ms condition. It was hypothesized that individuals that are good at speaker identification show a large difference in the thresholds for both conditions indicating a sharp auditory curve or a good frequency selectivity.

A center frequency of 2000 Hz was used for two reasons. First, the frequencies that contribute most to the intelligibility of speech lie in a higher frequency region (Elliott, 1963; Hirsch, 1952; Sher and Owen, 1974; Yoshioka and Thornton, 1980; Young and Gibbons, 1962). For example, French and Steinberg (1947) found that the most important frequencies for the over-all intelligibility of monosyllabic words lie in a range between 1500 and 2500 Hz. Actually, their study showed that when all frequencies above 1000 Hz are passed, the score is about 90%, whereas speech that contains only frequencies below 1000 Hz is only 27% intelligible. Second, concerning the topic of this research, higher frequencies (above 1000 Hz) appear to be more important for speaker identification than the lower frequencies (Compton, 1963).

The stimuli were generated by a Tucker Davis Technologies System 2 that consists of a 16-bit D-A converter, an anti-aliasing filter with a low-pass cutoff at 20 kHz, two programmable attenuators that attenuate in 0.1-dB steps from 1 to 99 dB and a

channel mixer. Subjects were seated in a sound-proof booth (Industrial Acoustics Company, Inc.) using TDH-50P headphones (Telephonics 29 6D 200-2).

3. Temporal resolution: Temporal resolution refers to the ability to detect changes in stimuli over time. For example, it refers to the ability to detect a brief gap between two stimuli or to detect that a sound is modulated in some way (Moore, 1997). As pointed out by Viemeister and Plack (1993), it is also important to distinguish the rapid pressure variations in a sound from the slower overall changes in the amplitude of those fluctuations. In other words, it is an indication of a person's resolution of changes in the spectral envelope of a signal: this ability should be crucial to speaker identification, because differences in voices result in different spectral shapes (Laver, 1980; Nolan, 1983). Temporal resolution also has been frequently shown to be related to speech perception (Glasberg and Moore, 1989; Irwin and McAuley, 1987; Tyler et al., 1982).

To measure temporal resolution a notched-noise procedure was used; specifically the subject was presented with two successive burst of noise where one of the two bursts was interrupted to produce a gap. The task of the subject was to indicate which burst contained the gap. In this study, a narrow-band masker was used together with a 10,000 Hz wide broadband noise that had a center frequency of 1000 Hz. The broadband noise was used to mask spectral splatter. That is, for narrow band noise, the introduction of a gap results in spectral splatter, which is energy spread outside the nominal bandwidth of the sound. This splatter would provide the listener with extra cues and therefore result in

inflated scores (Moore, 1997). In any case, both band noises were set at the intensity level of 50 dB, and the starting level for the gap size was set at 50 ms. The stimuli were generated by the cited Tucker Davis Technologies System 2 and subjects were seated in a sound-proof booth (Industrial Acoustics Company, Inc.) using TDH-50P headphones (Telephonics 29 6D 200-2). Each time the subject choose the correct answer, the gap size was decreased with 5 ms; however, when the subject made her first mistake, the gap was decreased/increased in 2 ms. steps. Each subject was confronted with 50 pairs of noise bursts. The thresholds were averaged from the first incorrect response to the end of the trial. The subject's score was based on the average of three trials and given in ms. In case of an outlier, only two scores were averaged.

### Musicality Assessment

The third series of auditory tests consisted of measurements of musical aptitude. It seems reasonable to assume that if earwitnesses use prosody as an identification clue (Hollien and Köster, 1996), their musical skills may be important to the process.

The Revised Seashore Test (1960) was used to test musical aptitude. The test was chosen, since it assesses, in addition to the main factors of pitch, tonal memory and rhythm, also timbre and intensity. It has been shown to be both valid and reliable (Horn and Stanov, 1982; Colwell, 1984) and even today, enjoys nearly the status of the standard test for individuals in primary schools, bands and other musical organizations. Indeed, most of its modules exhibit reliability coefficients greater than 0.70. Its administration

requires subjects to listen to various kinds of stimuli, such as pure tones, clicks, buzzes, and artificially synthesized complex tones. It consists of the following items: pitch, loudness, rhythm, time, tonal memory and timbre. These tests were administered with the subject placed in a sound-proof booth. Stimuli were played on a reel-to-reel (TEAC) coupled to Beyerdynamic headphones (DT211). Not all subtests were included as the parameter “time” was assessed by a test described in the auditory section. That is, time sensitivity was assessed by means of the gap detection test (testing temporal resolution). Those elements of the Seashore Test that were administered are as follows:

1. Pitch discrimination: This factor may be related to speaker identification for several reasons. First, auditors are able to perceive a speaker’s pitch and pitch variability and research on speaker identification has shown that those features can be used in this process (Compton, 1963; Ices, 1972; LaRiviera, 1971). Auditors are able to identify the speaker as a man (100-130Hz) or as a woman (190-220Hz) or as a child (300 Hz at 10 years of age), (Fant, 1956; Hirano, 1981). Within the spectrum produced by a particular gender, they also can specify whether a person speaks with an unusually high pitch or one that is higher (or lower) than someone else.

Subjects were presented fifty pairs of tones in the Seashore pitch discrimination test. In each pair, the listener had to determine whether the second tone was higher or lower in pitch than the first. The stimuli were generated by a beat-frequency oscillator through a circuit producing sinusoids. These tones were produced at about 500 Hz and



had a duration of 0.6 seconds. The initial five pairs had a frequency difference of 17 Hz after which the difference was slowly decreased to 2 Hz (for the last five pairs). A potential problem was observed during a pilot (procedural) study; subjects were confused by the instructions: should they compare the second tone to the first or vice versa? As with Davies (1978), an additional instruction was given that the subject should think of the direction of the tone: in other words, where is it going. Is the tone getting higher or lower? The approach was successful.

Each subject's number of correct answers was translated into a "percentile equivalent" -- a construct that indicates the proportion of the population who scored at or below the particular score. Norm tables were provided and used. For example, if a listener obtained a score of 43 on the pitch test, this value would be looked up in the table and also recorded; in this case it would be 60. That is, she scored as well or better than 60% of these adults tested during standardization.

2. Intensity discrimination: Even though the relationship between vocal intensity and speaker identification has not been investigated in detail, it is assumed that intensity level and variability are both useful as cues to speaker identification (Hollien and Köster, 1996). Speakers do differ in their level of speaking intensity and this parameter has been successfully used in automatic speaker recognition systems (Doddington, 1971; Lumnis, 1973; Rosenberg and Sambur, 1975). Further, in a related area, Scherer (1974) found that lay listeners primarily use pitch and loudness to judge vocal quality and Glasberg and Moore (1989) found that intensity discrimination is one of the factors that can predict a person's speech understanding ability.

To test intensity discrimination, the “loudness” submodule was used from the Revised Seashore test (1960). Fifty pairs of sinusoids were presented. The subject was asked to indicate for each pair whether the second tone is stronger or weaker than the first. The frequency was held constant at 440 Hz. The first five pairs started with a moderately large differential; one of 4.0 dB. The intensity difference was slowly made smaller until it had decreased to 0.5 dB for the last ten pairs. For scoring the results, see the preceding section.

3. Rhythmic discrimination: Since intonation and prosody are features that differ among speakers and can be both identified and remembered by listeners (Church and Schacter, 1984; Hollien and Köster, 1996), it was assumed that rhythmic discrimination would be related to speaker identification.

Accordingly, thirty pairs of rhythmic patterns were presented (Seashore Rhythm Test). Subject were required to indicate if the two patterns in each pair were the same or different. The source of these stimuli was a beat-frequency oscillator set at 500 Hz. and its tempo kept constant at the rate of 92 quarter notes per minute. The first ten items contained patterns of five notes in 2/4 time; the next ten, patterns of six notes in 3/4 time; and the last ten, patterns of seven notes in 4/4 time. The procedure for scoring was identical to the one used for pitch.

4. Timbre: Voice quality has been shown to be one of the most robust parameters in automatic speaker identification systems. The parameter, that is, the long term spectrum,

was found to show high accuracy levels, and is also resistant to the effects of speaker stress and to limited passband conditions (Bricker et al., 1971; Clarke and Becker, 1969; Doddington, 1970; Furui, 1978; Kosiel, 1973; Majewski and Hollien, 1974; Zalewski et al., 1975). Hence, an individual's sensitivity to voice quality or timbre (or tone quality) also was judged important.

The purpose of the timbre test is to measure a person's ability to discriminate between complex sounds which differ only in harmonic structure. It consisted of 50 pairs of tones; the subjects were required to judge if the tones in a pair were the same or different with respect to timbre or tone quality (Seashore Timbre Test). Each tone was made up of a fundamental component, with an F0 of 180 Hz and its first five harmonic overtones. Tonal structure was varied by reciprocal alteration in the intensities of the third and fourth harmonics with the alteration starting large and ending small. The scoring procedure was the same as that for pitch.

5. Tonal memory: This module tests how well listeners can remember a sequence of tones. Since intonation and prosody are features that differ among speakers (Darwin and Bethell-Fox, 1977) and that can be identified and remembered by listeners (Church and Schacter, 1984; Hollien and Köster, 1996), it was assumed that tonal memory would relate to identifying speakers.

This test included 30 pairs of tonal sequences consisting of ten items each of three- four-, and five-tone spans (Seashore Tonal Memory Test). In each pair, one note

was different between the two presentations, and the subject was to identify it. The 18 chromatic steps upwards from middle C were used; they were produced by a Hammond organ. Tempo was carefully controlled, and intensity was essentially constant. The scoring procedure was the same as that for pitch.

### Pilot Study

A pilot study was conducted with seven young females as subjects. The reason for doing so was that if test construction was appropriate, the procedure should produce a range of scores -- one that allows the formation of LOW-SPID and HIGH-SPID groups. The pilot study demonstrated that the setup of the identification test was appropriate for the task, as it showed a convenient range of scores was found; that is, subjects scored from 10% - 70% with values equally spread out over the continuum. In addition to the speaker identification selection procedure, all other tests cited in Table 2. were administered. However, no data of the pilot subjects were included in the main study. The goal of the practice runs was to familiarize the experimenter with the tests and the equipment and to ensure that the experimental devices were working properly.

## CHAPTER 3 RESULTS

### Introduction

This study was designed to investigate why some listeners are good at speaker identification while others are not. The approach used was the method of extremes whereby the characteristics of two groups are compared: a LOW-SPID group consisting of individuals who scored very low on a complex earwitness identification test and a HIGH-SPID group, consisting of listeners exhibiting very good scores. The criterion level for the superior group were scores of 55% and above, whereas a low score was defined as 10% or below. By using these thresholds two requirements were satisfied; the first was that the performance of the groups was reasonably different and, secondly, that the population was large enough for detecting differences between group means (i.e., power). Both groups were compared; those parameters that showed a significant difference in the means of the groups were considered important for identification. Three areas were investigated: the characteristics of the listeners pertaining to memory, psychoacoustic attributes and musical talent, since those were considered to be crucial for earwitness identification. Sample size was  $N=14$  for LOW-SPID and  $N=13$  for HIGH-SPID. As stated before, it was hypothesized that in all three areas the HIGH-SPID group would perform better than the individuals in the LOW-SPID group. This means that, according to this prediction, all values for the HIGH-SPID should be larger than the ones for the

LOW-SPID group, except for one of the psychoacoustic parameters. In the case of gap detection, or temporal resolution, a small score indicates good auditory sensitivity.

In order to choose the appropriate statistical tests, it was important to first study the distributions of the obtained data. If a normal or close to normal distribution can be assumed for each test, then parametric statistics can be applied whereas in the case of bimodal or skewed distributions, non-parametric tests should be used (Mendenhall 1990). Parametric tests would be preferred, since those are more powerful than non-parametric ones ( Marks, 1990). For example, if continuous response variables were treated as ordinal, a transfer necessary to perform non-parametric tests, would result in a loss of power. If it turns out that normal distributions can be assumed, then application of a t-test, for example, would be justified. To obtain a more detailed impression of the data and its distribution, the univariate SAS procedure was used. It provides several useful indicators: mean, standard deviation, median, mode and a measure of skewness. It also describes where the data are located (e.g., lowest 25%, highest 25%, etc.).

#### Memory Assessment

The first area to be researched is that of memory. The results can be found summarized in Table 3. It shows the means, standard deviations and differences plus an indication of whether the means support the predicted differences favoring the HIGH-SPID group (indicated with an asterisk). The statistical significance of these differences will be discussed later.

Table 3. Means and standard deviations for the memory tests. All values are in percentages, except for verbal memory.

| MEMORY TESTS:           | LOW-SPID |    | HIGH-SPID |    | Difference (%) | Trend |
|-------------------------|----------|----|-----------|----|----------------|-------|
|                         | Mean     | SD | Mean      | SD |                |       |
| Priming                 | 9        | 15 | 13        | 18 | 44             | *     |
| Verbal Memory           | 109      | 17 | 101       | 14 | -7             |       |
| Digit Span Forward      | 52       | 29 | 59        | 31 | 14             | *     |
| Digit Span Backward     | 63       | 25 | 75        | 24 | 19             | *     |
| Logical Memory I        | 69       | 31 | 60        | 25 | -13            |       |
| Logical Memory II       | 67       | 32 | 61        | 24 | -9             |       |
| Attention/Concentration | 44       | 24 | 56        | 28 | 27             | *     |
| Delayed Recall          | 69       | 19 | 66        | 14 | -4             |       |

\* Trends in the predicted direction.

Please note, that there are two parts to the table; first, priming and verbal memory are placed in their own section. Priming is not part of the Wechsler Test and its scores should be interpreted differently, especially since it is not expressed as a percentile equivalent. Also verbal memory results in an index score and is not expressed in percentage. The differences are defined as the percent increase from the mean for the LOW-SPID group. For example, the difference between the two groups for tonal memory is 29, which is 58% of 50, the mean for the LOW-SPID group.

First, are the data listed here close to what can be expected from normal subjects? The priming experiment was based on studies by Schacter and Church (1992, 1994) who studied voice-specific priming. Indeed, the priming value for voice reported for the student subject group (18-25 years) in Church and Schacter (1994, Experiment 1) was 13%; it is very similar to the values generated in this study, that is, 9% for LOW-SPID and 13% for HIGH-SPID. The slight difference could be caused by the difference in material used. The stimulus tape developed for this experiment was produced in the Phonetics Laboratory at the University of Florida. This means that the speakers on the present tape were different from the speakers used by Schacter and Church. Also, even though the word lists were the same, the selection of sections (same speaker, different speaker) may have been slightly biased. The scores for verbal memory, with a mean of 105, are higher than the score reported in the Wechsler documentation (1987) i.e. 100. The means of the next six tests are in percentile equivalent. This means that the average should be 50%. The Wechsler norms are based on normal subjects between 20 and 24 years old. Also here, the means are slightly higher, mostly around 60%. An explanation



for this phenomenon is the fact, that the Wechsler norms were based on the average population. However, all 112 subjects in this study were university students and it can be assumed that they have somewhat better memories than the average individual. Years of education correlate highly with all items of Wechsler Memory scores (Wechsler Memory Scale-Revised, 1987) with the scores improving with in increasing number of years of education. Education has been found to be highly correlated with intelligence also (Matarazzo, 1972).

Note that half of the eight memory tests show a difference in the means supporting the hypothesis (indicated with an asterisk \*) whereas half do not. Those of interest seem to be priming, both digit span tests and the attention/concentration module. For priming, the difference is largest, i.e. 44%; this parameter showed a score of 9% for the LOW-SPID group and 13% for the HIGH-SPID group. The attention/concentration module shows a 27% difference from 44% for LOW-SPID to 56% for HIGH-SPID. The ratios for the digit span tests are around 16%. When the standard deviations are considered, it may be observed that, they are again quite large (ranging from 14 index points for Verbal Memory to 31% for Logical Memory I). There apparently exists a wide variability in the data associated with the memory parameters. Considering the means in relation to the standard deviations, the following tests, at this stage, seem to be most meaningful in regard to the hypothesis: priming, both digit span backward and attention/concentration. They are the ones that show differences that are quite large and that are in the predicted direction.

## DIGIT SPAN BACKWARD

|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| HIGH-SPID | N        | 13     | 100% Max | 99 |
|           | Mean     | 75.2   | 75% Q3   | 90 |
|           | Std Dev  | 23.7   | 50% Med  | 82 |
|           | Skewness | -1.724 | 25% Q1   | 70 |
|           | Mode     | 70     | 0% Min   | 14 |
| LOW-SPID  | N        | 14     | 100% Max | 96 |
|           | Mean     | 63.4   | 75% Q3   | 90 |
|           | Std Dev  | 25.1   | 50% Med  | 70 |
|           | Skewness | -0.434 | 25% Q1   | 42 |
|           | Mode     | 70     | 0% Min   | 26 |

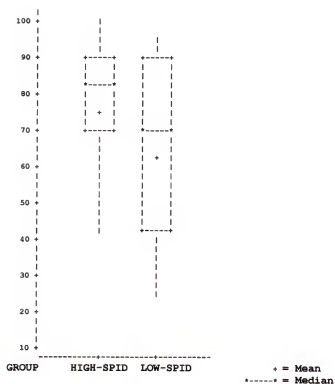


Fig. 2 Descriptive univariate SAS-plot of the digit span backward data.

The univariate statistic was used of SAS, a statistical analysis program, to obtain a clearer picture of the data obtained. This approach adds the median, mode, and range to the mean and standard deviation. It also provides an important measure of the normality of the curve, i.e. skewness. Figure 2. is such a display; one for digit span backward.

The means for both groups are indicated by a plus sign. As may be seen, Fig. 2 confirms the already observed large difference in the means between the HIGH-SPID and LOW-SPID groups. When observing the location of the means relative to the median, both distributions appear to be slightly asymmetrical; HIGH-SPID shows a mean that is lower than the median and the same is true for the LOW-SPID group. However, the most important measure here appears to be skewness. When a distribution is positively skewed, it has relatively few high scores and the pointed end is toward the right or positive direction. A negatively skewed distribution has the pointed end toward the left or negative direction. When skewness exceeds  $\pm 2.500$ , a normal distribution cannot be assumed for that particular data set. Both show a similar pattern: the distributions are slightly skewed toward the lower percentages (H skewness -1.724, L skewness -0.434). The LOW-SPID data set, however, shows quite a spread in the middle 50% of the data (from 42% to 90%, indicated by the rectangular box), resulting in a distribution that is rather extensive when compared to the sharper curve for the HIGH-SPID data. The size of the rectangular figure can be derived from the values for "75% Q3" (i.e. 90) and "25%" (i.e. 42%). The percentages indicate where that particular part of the data (e.g. first 25%) is located on the scale. The wide range results in overlapping distributions. Both curves are unimodal with a mode at 70. Overall, it can be shown that the distribution for digit

# ATTENTION/CONCENTRATION

|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 94 |
|            | Mean     | 56.4   | 75% Q3   | 82 |
|            | Std Dev  | 28.0   | 50% Med  | 65 |
|            | Skewness | -0.343 | 25% Q1   | 31 |
|            | Mode     | 82     | 0% Min   | 9  |
| LOW-SPID:  | N        | 14     | 100% Max | 87 |
|            | Mean     | 43.9   | 75% Q3   | 64 |
|            | Std Dev  | 24.0   | 50% Med  | 41 |
|            | Skewness | 0.197  | 25% Q1   | 22 |
|            | Mode     | 15     | 0% Min   | 9  |

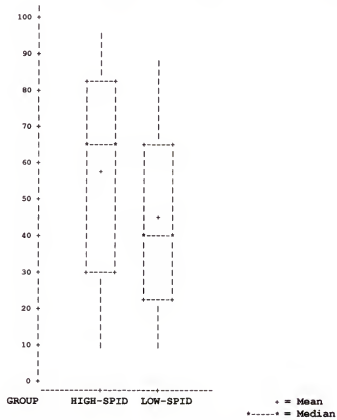


Fig. 3 Descriptive univariate SAS-plot of the attention/concentration data.

span backward can be assumed normal or close to it as skewness did not exceed 2.500, and the curve is unimodal. Therefore, parametric statistical tests for this parameter are justified.

The second univariate plot, that of the attention/concentration distributions, may be seen in Fig. 3. Here, the spread of both distributions is quite similar. Both are skewed, but in different directions: the HIGH-SPID distribution toward the higher percentages, the LOW-SPID toward the lower percentages (H skewness = -0.343, L skewness = 0.197). Even though the means are very different, the overlap of both curves is quite large. Also here, the distributions are unimodal. When Fig.2 and Fig.3 and the remaining plots (see Appendix C) were considered, it was concluded that all distributions associated with the memory tests can be assumed to be close to normal. Therefore, parametric tests also are justified for them.

At this juncture, it may be useful to calculate and study the p-values for the different distributions in order to determine whether the observed differences are statistically significant. Accordingly, a Two-sample One-Tail t-test was performed on each memory test. This procedure was used, because the hypotheses were all directional suggesting that the HIGH-SPID group would perform better than the LOW-SPID group in all tasks. The results of these tests can be found in Table 4. None of the differences between the means were statistically significant at the  $\alpha=0.05$  level. The pattern that was seen in Table 3. is of course repeated in Table 4: that is, both digit span backward and attention/concentration show the lowest p-values (i.e. 0.108 and 0.121 respectively), but they are not significant. Four of the eight tests showed a difference in favor of the

Table 4. Results of the Two-Sample One-Tail T-Test performed on the memory data.  
 Predicted relationship:  $\mu(\text{LOW-SPID}) < \mu(\text{HIGH-SPID})$

| MEMORY TESTS         | T       | DF   | p-value |
|----------------------|---------|------|---------|
| Priming:             | -0.6291 | 25.0 | 0.268   |
| Verbal Memory:       | 1.3287  | 25.0 | 0.902   |
| Digit Span Forward:  | -0.6063 | 25.0 | 0.275   |
| Digit Span Backw.:   | -1.2704 | 25.0 | 0.108   |
| Logical Memory I:    | 0.8263  | 25.0 | 0.792   |
| Logical Memory II:   | 0.5477  | 25.0 | 0.706   |
| Attention/Concentr.: | -1.1984 | 25.0 | 0.121   |
| Delayed Recall:      | 0.4640  | 25.0 | 0.677   |

Note: The P-values indicated with an asterisk (\*) are statistically significant at the  $\alpha=0.05$  level.

LOW-SPID group instead of the HIGH-SPID group; this is indicated by their high p-value. It means that these parameters have shown tendencies opposite to what the hypothesis predicted. For example, the verbal memory score for the LOW-SPID group was 109 and 101 for the HIGH-SPID group, the opposite of what was hypothesized. Therefore, the p-value is very high, i.e. 0.902. Overall, it can be concluded that, none of the memory parameters achieved significance. However, since this research was exploratory in nature and since the sample size was somewhat limited, note can be taken of certain trends and potential relationships. For example, the data for digit span backward and attention/concentration show reasonable tendencies and could be more rigorous predictors under other circumstances.

### Psychoacoustic Assessment

The second area of investigation in this study was the relationship between ability in speaker identification and the psychoacoustic characteristics of listeners. Three features were assessed: speech recognition (MRT in noise), temporal resolution (gap-detection), and frequency selectivity. These results may be found in Table 5. First, are the results to be found there similar to what can be expected for normal subjects? For the MRT, the scores were 65% and 68% for the LOW-SPID and HIGH-SPID groups respectively. These values are quite similar to the 72% reported by Kreul et al. (1968). A learning effect may explain why their value is slightly higher, since the listeners in Kreul's study took the MRT test (different forms) more than once. The listeners in this study were administered the MRT only once. Fatigue may be another reason for the slight difference in scores as, in this study, the MRT was always taken last in the session.

Table 5. Means and standard deviations for the psychoacoustic tests. The values for MRT are in percentages, the those for gap detection in ms. and the values for frequency selectivity in dB.

| TEST:             | LOW-SPID |       | HIGH-SPID |       | Difference (%) | Trend |
|-------------------|----------|-------|-----------|-------|----------------|-------|
|                   | Mean     | SD    | Mean      | SD    |                |       |
| MRT               | 65       | 11.6  | 68        | 10.7  | 5              | *     |
| Gap Detection     | 11.34    | 3.890 | 11.35     | 3.492 | 0.09           |       |
| Freq. Selectivity | 32.1     | 4.69  | 31.9      | 5.98  | 0.6            |       |

\* Trends in predicted direction.

Note also that the predicted relationship is:  $\mu(\text{LOW-SPID}) < \mu(\text{HIGH-SPID})$ , except for gap detection, where the reverse is predicted.

Second, the values obtained for gap detection were around 11 ms. for both LOW-SPID and HIGH-SPID. They are slightly higher and therefore slightly worse than the ones reported in a study where a similar setup was used (Shailer and Moore, 1983). In the latter study, they reported values around 8 ms. However, the results were based on only three subjects and, in addition, all subjects had experience or were familiar with the task. Since none of the subjects in this study had ever carried out the gap detection task before, it could be expected that they performed slightly worse.

In observation of the values obtained for frequency selectivity in this study they were found to be the same as those reported by Caffee (1997), who found a difference of



32 dB at 2000 Hz. As can be seen in table 5, the means for both the HIGH and LOW-SPID groups were 31.9 and 32.1 dB respectively.

It appears that one of the three psychoacoustic features tested shows a difference in favor of the predicted group, the HIGH-SPID group: the mean of the MRT scores for the HIGH-SPID group, 68% is 5% higher than the mean for the LOW-SPID group. However, the difference was small and the variability high (i.e., around 10% for both groups). Note that gap detection is the only characteristic where a negative correlation was predicted: it was hypothesized that a small value for gap detection would be associated with a good SPID score. However, there appears to be no difference between the experimental groups when gap detection was the criterion measure. The same appeared true for frequency selectivity, where both groups showed a score of 32 dB.

In consideration of the univariate plot of the MRT data (Fig. 4), two outliers can be noticed, indicated with an "0". Two subjects scored very high, 96% in the HIGH-SPID group and 100% in the LOW-SPID group. Since they created quite an abnormal interruption in the data flow, they were defined as outliers in the univariate graph. Both distributions are positively skewed (H skewness=1.530, L skewness=2.084) and these values are slightly higher than any of the others observed so far. Moreover, since the curves are unimodal (60 for HIGH-SPID and 70 for LOW-SPID) and do not exceed the skewness limit of 2.500, t-tests can be applied.

The results of the Two-Sample t-tests, performed on all psychoacoustic procedures may be found in Table 6. As stated earlier, the only p-value of interest is that

## MRT

|            |          |       |          |     |
|------------|----------|-------|----------|-----|
| HIGH-SPID: | N        | 13    | 100% Max | 96  |
|            | Mean     | 68    | 75% Q3   | 72  |
|            | Std Dev  | 10.7  | 50% Med  | 64  |
|            | Skewness | 1.530 | 25% Q1   | 62  |
|            | Mode     | 60    | 0% Min   | 54  |
| LOW-SPID:  | N        | 14    | 100% Max | 100 |
|            | Mean     | 65    | 75% Q3   | 70  |
|            | Std Dev  | 11.6  | 50% Med  | 63  |
|            | Skewness | 2.084 | 25% Q1   | 58  |
|            | Mode     | 70    | 0% Min   | 54  |

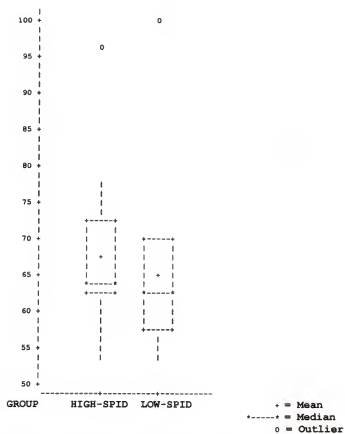


Fig. 4 Descriptive univariate SAS-plot of the MRT data.

Table 6. Results of the Two-Sample One-Tail T-Test performed on the psychoacoustic data. The predicted relationship was that  $\mu(\text{LOW-SPID}) < \mu(\text{HIGH-SPID})$  except for Gap Detection, where the relationship would be reversed.

| PSYCHOACOUSTIC TESTS: | T       | DF   | p-value |
|-----------------------|---------|------|---------|
| MRT                   | -0.6040 | 25.0 | 0.278   |
| Gap Detection         | -0.0119 | 25.0 | 0.988   |
| Freq. Selectivity:    | 0.0583  | 25.0 | 0.523   |

Note: The P-values indicated with an asterisk (\*) are statistically significant at the  $\alpha=0.05$  level.

for MRT, since group differences for gap detection and frequency selectivity were found to be the same or in the direction opposite from that predicted. However, significance was not found for MRT either.

### Musicality Assessment

The results of the music data can be found summarized in summary Table 7. Specifically, the LOW-SPID cohort is compared to the HIGH-SPID. First, what scores can be expected based on other studies? All parameters in Table 7. are part of the Seashore Test (1960) and therefore all means are expressed in percentile equivalent. The Seashore equivalents were based on at least 4000 students in grades 9-16 per subtest. The values represent the percentage of the population that scored at that level or lower. Therefore, if the score is over 75, they would rank in the upper quadrant; if 25 or lower in the lower quadrant. Theoretically, the means should all be around 50%. Both tonal

memory and loudness fall around the 50% level, with pitch, rhythm and timbre deviant. Rhythm was significantly higher with both values around 75%, but the scores for timbre were around 30%. It is unclear why these differences exist. Rhythm, however, might have been higher, because of the elaborate explanation and the examples given in this study. The quality of the tape might have explained the lower scores for timbre. A moderate amount of background noise may have affected timbre more than other tests. The score for pitch was 41% slightly below 50%. Of course, the small sample size may be another reason for the discrepancies.

Table 7. Means and standard deviations for the music tests. All values are in percentages.

| MUSICAL TEST: | LOW-SPID |      | HIGH-SPID |      | Difference (%) | Trend |
|---------------|----------|------|-----------|------|----------------|-------|
|               | Mean     | SD   | Mean      | SD   |                |       |
| Pitch         | 41       | 29.9 | 41        | 23.2 | 0              |       |
| Loudness      | 46       | 22.6 | 47        | 23.4 | 2              | *     |
| Rhythm        | 73       | 27.1 | 79        | 19.5 | 8              | *     |
| Timbre        | 28       | 21.4 | 33        | 21.1 | 19             | *     |
| Tonal Memory  | 50       | 25.4 | 79        | 22.4 | 58             | *     |

\* Trends in predicted direction.

Of the five tests four show a difference in a direction suggesting that the HIGH-SPID subjects are better at the tasks than the LOW-SPID ones. However, can it be concluded from the data, that people that have talent for music are also good at identifying speakers? In observation of the individual tests, it can be seen that not all differences are equally large. For loudness the means for both groups are nearly identical, resulting in a 2% increase. Both rhythm and timbre show moderate differences that are slightly larger, 7% and 15% respectively. It can be seen that the one for tonal memory is the most convincing of all, being 58%. Overall, it can be concluded that three of the five tests show a difference favoring the HIGH-SPID group which also agrees with the hypothesis expressed earlier: rhythm and timbre showing only a moderate difference and tonal memory showing the largest difference. However, can it be concluded that these differences are meaningful? By consideration of the standard deviations, it may be observed that these are rather large, varying from 20% to 30%. This means, that the data show a great variability and that the trends may be weak.

Figures 5. and 6. display the results of the univariate procedure for pitch and for tonal memory. As may be seen, the "pitch" means for both HIGH-SPID and LOW-SPID are the same. When observing the location of the means relative to the median, both distributions appear to be slightly asymmetrical; HIGH-SPID shows a mean that is lower than the median and the opposite is true for the LOW-SPID group. However, the most important measure here appears to be skewness. For pitch, both values for skewness indicate positively skewed distributions (HIGH-SPID: skewness = 0.059, LOW-SPID: skewness = 0.707): for both groups the top 50% is spread out over a larger area than the

## PITCH

|           |          |       |            |    |
|-----------|----------|-------|------------|----|
| HIGH-SPID | N        | 13    | 100% Max   | 84 |
|           | Mean     | 40.8  | 75% Q3     | 52 |
|           | Std Dev  | 23.2  | 50% Median | 45 |
|           | Skewness | 0.059 | 25% Q1     | 23 |
|           | Mode     | 45    | 0% Min     | 5  |
| LOW-SPID  | N        | 14    | 100% Max   | 92 |
|           | Mean     | 41.2  | 75% Q3     | 68 |
|           | Std Dev  | 29.9  | 50% Med    | 29 |
|           | Skewness | 0.707 | 25% Q1     | 20 |
|           | Mode     | 20    | 0% Min     | 6  |

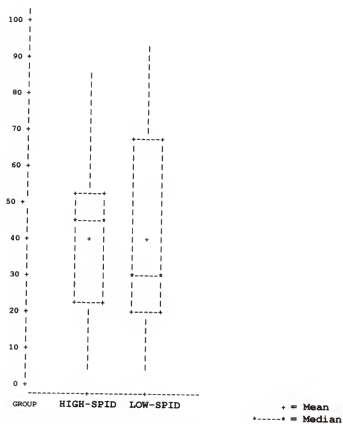


Fig. 5. Descriptive univariate SAS-plot of the pitch data.

bottom 50% and therefore the pointed end is toward the higher values. From the figure, it also becomes clear why the standard deviations are so large for both groups (i.e. H: 23% and L: 30%). First, the spread or range of both data sets is quite large (H: 79% and L: 86%) and in addition, the middle 50% of the data (again, indicated by the rectangular box) are extremely spread. The mode is 45 for HIGH-SPID and 20 for LOW-SPID. Both values indicate a unimodal distribution. Overall, it can be shown that the distribution for pitch can be assumed normal or close to it as skewness did not exceed 2.500, and that the curve is unimodal. Therefore, parametric statistical tests for pitch are justified.

Second, the distribution of the musical test that showed the largest difference was tonal memory. Fig. 6 shows means that are quite different from each other, they fall at the 79% level for the HIGH-SPID group and at 50% for the LOW-SPID group. Moreover, the overlap in this instance is much smaller than that for pitch. Here, both distributions are slightly skewed toward the lower values (H skewness= -0.989, L skewness= -0.761), but the degree of skewness does not exceed the limit of 2.500. A ceiling effect can be noticed for the HIGH group: the middle 50% of the data are in the upper part of the range overlapping with the upper 25%. It means that for this group, the tonal memory test may not have been challenging enough. When looking at the ranges, it may be seen that they are quite extensive, a relationship which explains the large standard deviations. The modes are 99 for HIGH-SPID and 52 for LOW-SPID. Also after considering the distribution for tonal memory, unimodal and only slightly skewed, a parametric statistical test can be justified.

## TONAL MEMORY

|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| HIGH-SPID | N        | 13     | 100% Max | 99 |
|           | Mean     | 78.8   | 75% Q3   | 99 |
|           | Std Dev  | 22.4   | 50% Med  | 85 |
|           | Skewness | -0.989 | 25% Q1   | 61 |
|           | Mode     | 99     | 0% Min   | 29 |

|  |          |        |          |    |
|--|----------|--------|----------|----|
|  | N        | 14     | 100% Max | 85 |
|  | Mean     | 50.0   | 75% Q3   | 72 |
|  | Std Dev  | 25.4   | 50% Med  | 57 |
|  | Skewness | -0.761 | 25% Q1   | 29 |
|  | Mode     | 52     | 0% Min   | 1  |

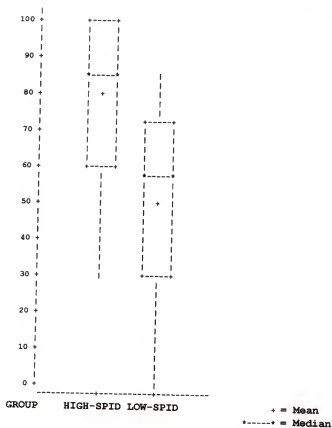


Fig. 6. Descriptive univariate SAS-plot of the tonal memory data.



By observation of the remaining plots (See Appendix C), it may be seen that they are similar to, but fall between, the ones described above. All exhibit large standard deviations and many distributions are slightly skewed, (however, none exceed the skewness>2.500 level). To be specific, the plot for loudness appeared very similar to that for pitch. Both rhythm and timbre showed means that are quite different from each other, but had overlapping distributions due to extensive variability. In any case, it was concluded that a parametric approach was justified for all of these factors. Accordingly, a parametric two-sample one-tail t-test was performed on each of them. The results of these tests can be found in Table 8.

Table 8. Results of the Two-Sample One-Tail T-Test performed on the music data.  
Predicted relationship:  $\mu(\text{LOW-SPID}) < \mu(\text{HIGH-SPID})$

| MUSIC TESTS | T       | DF   | p-value |
|-------------|---------|------|---------|
| Pitch:      | 0.0386  | 25.0 | 0.515   |
| Loudness :  | -0.0904 | 25.0 | 0.464   |
| Rhythm:     | -0.6448 | 25.0 | 0.263   |
| Timbre:     | -0.6351 | 25.0 | 0.266   |
| Tonal Mem:  | -3.1147 | 25.0 | 0.002 * |

Note: The P-values indicated with an asterisk (\*) are statistically significant at the alpha=0.05 level.

As may be seen, the large variability seemed to interfere with significance regarding the difference in the means. Hence, only one significant difference was found, that is, the one for tonal memory ( $p=0.002$ ). Note, that when looking at the P-values, the pattern of Table 7. with the means is roughly the same as the one in Table 8. Pitch and loudness with p-values close to 0.500, appear to be at one end of the spectrum with tonal memory with a significant p-value of 0.002 at the other end. Overall, only tonal memory appears to show a statistically significant difference favoring the predicted direction.

### Summary of the Results

In conclusion, it may be seen that (of the three areas investigated) a few relationships showed slight trends, but only one was statistically significant. That is, differences for the music parameter of tonal memory were in the predicted direction and statistical significance was reached. When the memory parameters were assessed, only digit span backward and attention/concentration showed a trend -- but neither was significant. None of the three psychoacoustic parameters studied exhibited any correlation with talent in earwitness identification.

At this juncture, it was considered useful to see whether the pattern of the means remained consistent when the groups were decreased in size. Hence, the means were calculated for the LOW-SPID and HIGH-SPID groups that were smaller ( $N=7$ ) and at even more extreme ends of the spectrum: that is, the groups consisted of listeners with, respectively, a score of 5% and below and of 60% and higher. The pattern that was found

confirmed the one for the larger groups for all parameters except for only one; the trend for verbal memory changed in the opposite direction. Moreover, in more than half of the cases, the differences were more robust than for the larger groups. From this analysis, it was concluded that the pattern of means observed for the HIGH and LOW-SPID groups is a moderately stable one.

### Fitting a Model

This research does not appear to explain why some individuals are adept at recognizing speakers from their voices and others are not. Nevertheless, certain possible relationships were observed. Tonal memory was a robust predictor but since this research was exploratory in nature and since the sample size was relatively small, it does not seem to be justified to conclude that only this factor is of importance. Other parameters may very well be useful -- and so demonstrated by future studies employing a large sample size and perhaps a modified research design.

The t-tests that were performed showed which parameters may be important for speaker identification. However, they were tested by themselves; in other words, nothing is known about the way they interact with other parameters. Moreover, it is also not clear how they can predict the validity of a witness. When those questions are answered, a tentative model can be developed. It is quite possible that a reasonably good predictive model can be developed from the present data. One was tried; it is based on attempts to discover and describe those patterns which actually exist in the data set of interest. That is, a model may be established that can predict the dependent parameter (y) from the

independent parameter  $x$  by defining observed patterns and estimating appropriate parameters. Since categorical data were present in this study (i.e., HIGH-SPID and LOW-SPID), the linear parametric logistic regression approach had to be applied (Hosmer and Lemeshow, 1989; Agresti 1984). Specifically, the Logistic SAS-procedure was used. It fits linear regression models for binary data by the method of maximum likelihood. First, this procedure was used in a “stepwise” manner: this means that separate multiple regressions are sequentially performed on each parameter. The effect of each is calculated on the validity of the construct: the model is tried with and without each parameter and if it turns out that one has made a significant contribution ( $\alpha=0.05$ ), it is defined as a good predictor. When the stepwise procedure was applied, only tonal memory exhibited significance. One of the major disadvantages of this approach, however, is that it does not allow control over the variables. The experimenter cannot select the parameters for the model, since they are all selected by default. Moreover, the procedure only provides those that were significant, but does not give the p-values themselves. Therefore, linear multiple regression was employed again, but this time using the procedure whereby the covariates could be selected manually. Theoretically, the model could be repeatedly tried with different combinations of the 16 parameters. However, to be time efficient, the total number of features was reduced to those which demonstrated at least some kind of a trend in favor of the HIGH-SPID group. That is, only those tests with a t-test p-value below 0.300 were considered at all. Table 9. provides the parameters/tests that satisfied this requirement.

The second set of relationships studied were the interactions. In other words, some

parameters might not be good predictors on their own but would predict quite well when coupled with another parameter. However, before any attempt was made to fit these combinations into a model, the question was asked as to whether it could be simplified in any way. That is, those parameters should not be included that mask the strength of the effect of the stronger variables. For example, redundant variables should be avoided and parameters which correlate highly with each other should not both be entered into the model. When parameters correlate with each other, it means that they are co-linear; when one increases, the other shows the same increase. In order to avoid co-linearity in the model, a correlation matrix was calculated using the CORR SAS procedure. It shows the correlation between the tests cited in Table 9, for both the LOW-SPID and HIGH-SPID group (See Appendix D).

Table 9. Tests that satisfy the requirement for entering the model.

| TESTS:                  | p-value |
|-------------------------|---------|
| Tonal Memory            | 0.002   |
| Digit Span Backward     | 0.108   |
| Attention/Concentration | 0.121   |
| Rhythm                  | 0.263   |
| Timbre                  | 0.266   |
| Priming                 | 0.268   |
| Digit Span Forward      | 0.275   |
| MRT                     | 0.276   |

Note: The requirement is  $p \leq 0.300$ .

Note that the values in the table should be considered as they are. As will be seen, most of the tests show a low correlation with the others. For example, when the coefficients for the HIGH-SPID group are considered, the one between tonal memory and timbre is only 0.043. This suggests that if rhythm changes with one unit (i.e., 1.00), tonal memory only increases 0.043. In other words, hardly any correlation exists between the two parameters. However, digit span backward and attention/concentration appeared to be parameters that are highly correlated ( $r=0.84$ ). The next coefficient in decreasing order was 0.75 for digit span forward and attention/concentration. At this juncture, it can be asked if this pattern also could be found for the LOW-SPID group. As it turned out, the same pattern was found only for one of the relationships just described. The digit span forward / attention concentration coefficient remained about the same, at 0.78. The one for digit span backward / attention concentration dropped however to 0.62. Therefore, only the tests with a consistent high correlation for both groups were considered. It was decided that only one of the two parameters, digit span forward and attention/concentration, should be employed in support of the model. Attention/concentration was chosen, since it showed the stronger relationship with speaker identification. The remaining seven parameters were used to construct the model.

The results of the interaction analysis may be found in Appendix E. When the p-values are considered, it can be seen that none of the interactions are significant. However, it was decided to keep two interactions, since it is possible that, even though non significant, they still can contribute positively to a model. Tonal memory and

attention/concentration ( $p=0.156$ ) and tonal memory and MRT ( $p=0.179$ ) were the two that met these criteria. Accordingly, the model was tried including the seven main factors discussed earlier and the two interaction terms. The results can be found in Table 10.

Table 10. Estimates and p-values of the first model

| Combination           | Parameter Estimate | p-value |
|-----------------------|--------------------|---------|
| Rhythm                | 0.2470             | 0.6300  |
| Timbre                | -0.1588            | 0.7309  |
| Tonal                 | 0.5893             | 0.7871  |
| Priming               | 49.8537            | 0.5810  |
| DigitSpanB            | -0.0952            | 0.7011  |
| Attention/Conc.       | 1.6474             | 0.5360  |
| MRT                   | -1.4423            | 0.5356  |
| Tonal Mem*Attent/Conc | -0.0211            | 0.5125  |
| Tonal Mem/MRT         | 0.0188             | 0.4460  |

| Criterion | Intercept Only | Intercept and Covariates | Chi-Square for Covariates       |
|-----------|----------------|--------------------------|---------------------------------|
| AIC       | 39.393         | 30.224                   | .                               |
| SC        | 40.689         | 43.183                   | .                               |
| -2 LOG L  | 37.393         | 10.224                   | 27.169 with 9 DF ( $p=0.0013$ ) |
| Score     | .              | .                        | 12.394 with 9 DF ( $p=0.1920$ ) |

Note: the p-value is the p-value for this entire model

The p-value for the entire model ( $p=0.001$ ) indicates that it is a very good fit.

However, when the p-values for the individual terms are assessed, those for tonal memory

( $p=0.787$ ) and timbre ( $p=0.731$ ) appear quite large. Thus, it was possible that those individual  $p$ -values could be improved by fitting a model with a different combination of terms. By looking at the original  $p$ -values for the  $t$ -tests listed in Table 9., it was decided that MRT be eliminated. Of all the relationships, this one showed the weakest relationship. Removal of MRT would result in a model consisting of factors with  $p$ -values of only 0.268 (priming) or below. The model was reconstructed and, this time without MRT or its interaction. The results may be seen in Appendix F. As it turned out, this procedure improved the model with a change in  $p$ -value from 0.001 to 0.0006. Also, the individual  $p$ -values decreased sharply. In the first model, the average  $p$ -value was around 0.580 whereas in this model they are roughly about 0.150. The highest score was found to be for digit span backward (0.274). Since this one was quite large compared to the remaining values, it was reasoned that the model would probably also benefit from removing this term. A final attempt was made to improve the  $p$ -values by removing digit span backward. The results of this effort can be found in Table 11. As may be seen, the overall  $p$ -value for the entire model shows that this combination provides a very good fit:  $p=0.0006$ . Secondly, the  $p$ -values of the individual scores are seen to improve significantly with tonal memory, attention/concentration and the interaction exhibiting values of about 0.075. Additionally, the maximum value obtained in this case was only 0.208 for timbre. The model now appeared at an optimum level and no more attempts were made to improve it. It may be seen displayed in Fig. 7. It includes the participating parameters in order of significance starting with the term with the best  $p$ -value of 0.074, that is, attention/concentration. From left to right the  $p$ -values increase slowly and end with the least significant value of  $p=0.208$  for timbre.



Table 11. Estimates and p-values of the final model.

| Combination           | Parameter Estimate | p-value |
|-----------------------|--------------------|---------|
| Rhythm                | 0.0984             | 0.153   |
| Timbre                | -0.0847            | 0.208   |
| Tonal                 | 0.9507             | 0.078   |
| Priming               | 22.4989            | 0.158   |
| Attention/Conc.       | 0.8381             | 0.074   |
| Tonal Mem*Attent/Conc | -0.0109            | 0.077   |

| Criterion | Intercept      |                          | Chi-Square for Covariates       |
|-----------|----------------|--------------------------|---------------------------------|
|           | Intercept Only | Intercept and Covariates |                                 |
| AIC       | 39.393         | 27.661                   | .                               |
| SC        | 40.689         | 36.732                   | .                               |
| -2 LOG L  | 37.393         | 13.661                   | 23.732 with 6 DF ( $p=0.0006$ ) |
| Score     | .              | .                        | 11.115 with 6 DF ( $p=0.0849$ ) |

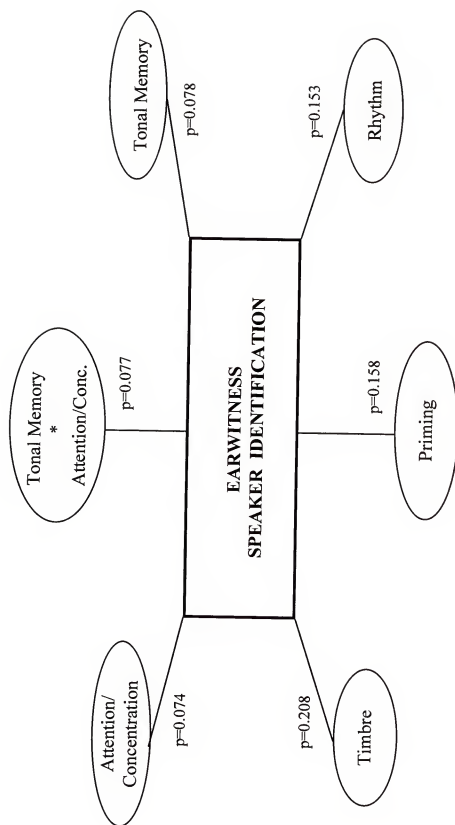
Note: the p-value is the p-value for this entire model

Note, that only in this combination do the parameters behave this way. That means that removing one term will affect the whole model and all p-values will change. Therefore, all scores have to be interpreted within this model.

In summary, even though the parameters were not significant on their own, their combined contributions have resulted in a model that is of some predictive value. Also, some parameters performed better in the presence of others -- an additional reason as to why the overall model improved as it did. Note, however, that the model is mainly used for descriptive purposes and the model should not be considered as the "perfect" fit.

Linear regression was employed here, but it may well be that non-linear regression would have resulted in an even better fit. Also, only part of the parameters were entered and tried. However, trying all possible combinations of factors in the model or attempting non-linear regression methods was considered beyond the scope of this research.

Again, tonal memory is the one parameter that this study has shown to be most important. However, with the construction of the model, other parameters are provided that may serve as indicators of an earwitness' ability to identify speakers. At least they may be considered for further study.



P-value of entire model: 0.0006

Fig. 7 A speaker identification model for assessing earwitnesses

## CHAPTER 4 DISCUSSION AND CONCLUSION

### Introduction

This study was carried out to investigate the relationships between the earwitnesses who were superior and inferior at speaker identification accuracy. It was hypothesized that speaker identification shows a positive correlation with a listener's 1) memory skills, 2) auditory function and 3) musicality.

### Discussion of the Results of the Assessment of Memory

It was hypothesized that good memory is positively correlated with excellence in speaker identification. In this case, four of the parameters showed differences in favor of the HIGH-SPID group. This group obtained better scores for priming, digit span forward and backward, and attention/concentration than the LOW-SPID group. However, due primarily to the large standard deviations, none were significant. Of these the attention/concentration attribute appeared to correlate best with identification.

Even though priming appeared to contribute positively when entered with the other factors in the model, it was not significant when a t-test was performed. Why this is the case, is unclear. Errors in design and administration could have been one reason. That

is, the instructions to subjects might not have been clear enough at the initial stages. However, since the difference between the groups was 44% and since the parameter contributed positively to the model, it should not be ignored in future research. Since this experiment was a replication of Schacter et al. (1994, Experiment 1), it can be assumed that the priming score is an indication of implicit memory and of implicit memory only. Therefore, the data can be used as an indication that the implicit or unconscious memory of a speaker's voice may influence their ability to identify speakers. A voice may be unconsciously familiar and therefore evoke a first reaction of recognition without the listener knowing why (Jacobi et al, 1993). This reaction can be considered to be an automatic response in contrast to conscious or intentional recognition (Jacoby et al, in press). The priming data suggest that implicit recollection of a talker's voice characteristics contribute to speaker identification accuracy. So, even if witnesses cannot explain why they have identified a certain talker as the target, their choice should still be considered valid; they may have been guided by implicit memory. In short, priming should not be ignored as a factor in testing for earwitness validity.

Another attribute that was not significant but which showed a slight positive trend was digit span backward: the difference between the means was almost 20%. Since, the limits of a person's STM affects the way information is stored (Schmajuk and Dicarlo, 1991), it was hypothesized that it could influence voice-specific memory. However, this relationship was not confirmed. Also Digit span forward is an indication of a person's STM and that parameter did not seem to contribute either. At this juncture, it cannot be argued that STM affects speaker identification.

Attention/concentration is another parameter that showed quite a difference (i.e., 27%) between the means of the HIGH- and LOW-SPID group and that despite of being non-significant in the t-test, contributed to the model both in an interaction and by itself. The score here consists of two other scores: digit span backward and mental control. The resulting value is an indication of the distractibility of a person. That is, it measures how well he/she can concentrate. It was hypothesized that concentration skills relate to speaker identification and a trend was found that seems to confirm this hypothesis. This finding agrees with the positive correlation that exists between attention and memory (Alain and Woods, 1997; Kellog et al, 1996, Mulligan and Hartman, 1996; Norman, 1976; Schmitter-Edgecombe, M. 1996). The positive trend found in this study also indirectly suggests that when a victim is aroused, due perhaps to life threatening circumstances, it will improve memorization of the event. Of course, this is only true if arousal increases a person's attention or concentration. Biological studies on the effects of threat, however, seem to justify this assumption (Magoun, 1963; Whybrow, 1994 ). The positive correlation between stress and memory is supported by the results of study by Atwood and Hollien (1986) and indirectly by the Rusted and Dighton (1991) and Mealey et al. (1996) research. Atwood and Hollien found that aroused subjects were significantly better at speaker identification than subjects who were not stressed. Rusted and Dighton (1991) found that when people were confronted with a picture of someone that was said to be a "cheater", his/her face was better remembered than the faces of individuals that were depicted as having an honest or more positive character. Mealey et al. (1996), studying spider phobics suggest that information that pertains to threatening objects is better

retained than neutral information. Further, Christianson and Engelberg (1997) suggest that there are biological reasons for the existence of this phenomenon. That is, the detailed memories of a threatening situation may evoke warning signals in the case of similar dangerous situations experienced in the future. Being able to recognize threatening situations may be crucial for survival. Robbins (1997) defines the cholinergic system as that part of the brain responsible for memory enhancement under circumstances of arousal. However, it also appears that another seemingly contradictory process exists. Sometimes, it may be more important to forget memories in cases of extremely traumatic experiences (Christianson and Engelberg, 1997). "Forgetting" here means, "not being able to retrieve"; the memories of the experience still exist, but due to repression, it may take time to recall them and sometimes they can never be recalled or only incorrectly (Brewin, 1997; Conway, 1995; Christianson and Engelberg, 1997; Roe and Schwartz, 1996). Since both processes apply to the memory of earwitnesses, it is important that they be acknowledged and recognized by scientists and personnel involved with witnesses.

Of course, even though the data supported a positive trend, one factor cannot be ruled out and that is a common problem in research where subjects are evaluated; that is, here listeners in the LOW-SPID group may have performed poorly at both the identification and the attention/concentration tasks due to lack of motivation. In other words, since they may not have been really interested in participation, they did not do well in either task regardless of their real memory ability. Since most subjects seemed to enjoy the experiment, it is not assumed that lack of motivation was a primary cause. However, it cannot be ruled out.

In summary, it seems that their attention/concentration skills explained why the listeners of the HIGH-SPID group did so well at speaker identification. Being able to focus and concentrate on a certain stimulus while ignoring signals that are less important improves memory of the stimulus of interest. The judgment of earwitnesses that were highly concentrated on the crime, can be predicted to be more valid than the one of witnesses that were not really involved. Second, the individual attention/concentration skills of a person should also be considered for inclusion in earwitness assessment tests.

It seems that learning and retention skills of verbal material do not play a role in speaker identification. Word-specific memory may not serve as a retrieval cue for voice recognition. In other words, being able to repeat the verbal stimulus, may not invoke or strengthen voice related memories that could be useful even if a different verbal stimulus is given during the lineup. All factors which did not exhibit a positive correlation, are associated learning and immediate and delayed recall of verbal material: Logical memory I and II, verbal memory and delayed recall. Of course, the conclusion stated above only holds if the listeners in this experimental set up were actually attempting to remember the monologues of the confrontation tape verbatim. However, since the initial instructions for the confrontation with the four target speakers were very similar to the instructions used for the verbal memory test, it seems that the assumption is justifiable. In both cases, the subjects were asked to try to remember as much as they could of the words and content of the monologues of each speaker. As stated earlier, this approach was used to avoid that they would explicitly focus on the voice of the speaker. Whether they really did



remember anything of the monologues was not measured. It would be a useful parameter to study in future research. Another parameter that showed support for the assumption above is explicit memory assessed in the cued recall test of the priming experiment. Apart from the stem completion test where subjects were required to write down the first item that came to mind, the same test was repeated but with different instructions. In the cued recall test, the task was to write down the items that were remembered from the initial meaning test. The score is believed to indicate conscious or explicit recollection (Schacter and Church, 1992; Schacter et al. 1994). Considering the results pertaining to verbal memory discussed above, one would not expect to find a relationship between explicit memory and speaker identification. Indeed, the score for explicit voice memory for the LOW-SPID was 0.20 and for HIGH-SPID 0.16. It indicates that conscious memory about a word-speaker combination does not affect speaker identification. One explanation for verbal memory being non significant is that listeners that concentrate entirely on the content of the message, might not also be able to concentrate on anything else. That means that they might remember but little of the talker's voice. Since the hypotheses were all directional (i.e., one tail), verbal memory was not considered for inclusion in the model. However, it probably would be useful to investigate those relationships in future research.

In summary, the data suggest that remembering what the speaker said during the crime, probably does not reinforce memory of the voice-specific features underlying the message. Also, a good memory for content may indicate that the witness was not able to pay the necessary amount of attention to the voice.

Even though none of the memory factors tested proved significant, it is still assumed that memory is important for eyewitness identification. An explanation for the lack of it may be that the sample size was too small for studying abilities that are this complex. Another reason may have been that the Wechsler Memory Scale is too focused on pathology and is not sensitive enough to properly assess differences in normal individuals. Perhaps studies with different memory tests, or with adjusted ones, might exhibit means with smaller standard deviations and ones that are significantly different. At least, this study, can serve as a guide for research of that type.

#### Discussion of the Results of the Psychoacoustic Assessment

Three psychoacoustic measures were obtained: speech recognition in noise (MRT), temporal resolution (gap detection), and frequency selectivity. No significant difference was found. Again, this may be due to the small sample size. The means for temporal resolution and frequency selectivity were the same; hence, it may well be that only a normal or close to normal skill level here is required for identifying a speaker.

The fact that the MRT did not show significance is a little surprising. Test scores of this type are indications of how well a person can select or filter out the verbal stimulus from competing stimuli. The same skills also seemed crucial to speaker identification. After the crime has been committed, an eyewitness' speech perception abilities are the only source of information he/she can rely on in the voice lineup. In other words, the results of the witness' speech filtering process, serve as input for other important processes that occur later in time (e.g. analysis of the signal for storage). Thus, one of the

reasons why the MRT did not show to be significant may be that the construction of the test was not ideal. Perhaps the speech stimuli were not sensitive enough. The selection of items for the multiple choice test may not have been optimal. Using a different type of masking (e.g., speech noise, multi-talker babble) may yield different and may be more useful results.

### Discussion of the Results of the Assessment of Musicality

Five tests were drawn from the Seashore Musical test to assess individuals aptitude. As was seen, three of the five music parameters showed differences in means that varied from 10% for rhythm to almost 60% for tonal memory. However, only the last parameter was statistically significant but all three exhibited positive trends and, hence, ultimately were able to contribute in the model. No correlation with identification was observed for pitch and loudness. So, can it be concluded that the results pertaining to musicality support the hypothesis that there exists a positive correlation between musical aptitude and speaker identification accuracy? Is it true, that people with a good musical aptitude also do well in speaker identification? The fact that three factors tested exhibited a positive trend, does argue that the hypothesized relationship exists. Also, when an overall mean was calculated of all musical test scores for both the HIGH-SPID and the LOW-SPID group, the difference between them was quite apparent; the means were 56% and 48% respectively. In addition, the p-value ( $p=0.072$ ) of the two-sample t-test suggested at least the existence of a (non significant) trend, where speaker identification is facilitated by musical aptitude.

Of all parameters, tonal memory showed to be the only parameter with significantly different means for the two groups ( $p=0.002$ ). The Seashore test is an assessment of how well an individual can remember a sequence of different tones. Therefore, it was assumed that this skill should be related to a listener's perception of shifts in speaking fundamental frequency of a speaker, that is intonation. Considering the significant p-value, the data of this study seem to support the assumption. The other parameter related to the perception of a speaker's intonation or prosody is rhythm. The rhythm module is assumed to be related to a listener's perception of the temporal shifts in a person's speech. Although not significant, it appeared to positively contribute to the model. The fact that both rhythm and tonal memory seem to affect identification agrees with the findings of earlier studies that intonation and prosody can be identified and remembered (Church and Schacter, 1984; Hollien and Köster, 1996) and that speakers differ in intonation (Darwin and Bethell-Fox, 1977). Of course, a high sensitivity to intonation and prosody of a speaker would be even more useful, when, during the first confrontation with the speaker and during the lineup, the type of speech used is the same. For example, when the speech in the lineup is extemporaneous, it would facilitate the comparison of the samples produced by the suspect and foils with the speech that was heard at the time of the crime. However, even though, extemporaneous versus read speech was used in the lineups in this study, skills related to perceiving intonation and prosody still would seem very useful. Read speech was used in the lineups of this study, since that is closest to real world cases. Obtaining extemporaneous speech for the voice parades, even though preferred (Broeders, 1996), is quite often a much more complicated

procedure than using verbatim speech from a transcript (Laubstein, 1997).

The third music parameter that showed a slight trend was timbre. The difference in the means was almost 20% with the higher mean (i.e., 33%) for HIGH-SPID. Due to the high variability of the data, the t-test did not result in a significant p-value. However, the trend was a little difficult to ignore. Moreover, when tried as a parameter in the model, it showed a positive contribution. Due to a moderate amount of background noise, the sensitivity of timbre test may have been slightly reduced. However, the fact that it contributed to the earwitness speaker identification model suggests that it should be considered for future research. Timbre or quality discrimination skills of tones also may relate to timbre discrimination skills for human voices.

The factors that did not exhibit any correlation were pitch and loudness; in this instance, the scores for both LOW-SPID and HIGH-SPID groups were practically the same. In regard to pitch discrimination, the data seem to suggest that although people are able to, for example, perceive pitch, it does not affect speaker identification. This contradicts the findings of others who found that listeners do use pitch as a cue for that task (Compton, 1963; Ices, 1972; LaRiviera, 1971). The same suggestion seems to hold for loudness: this parameter also did not influence identification scores. Perhaps it is true that, in order to successfully carry out an identification task, a listener does not need to excel in pitch and loudness perception, but just needs to have a normal or close to normal sensitivity. The cognitive process that occurs after pitch and loudness have been identified may influence speaker identification accuracy more than the basic skills of, for example, pitch definition.

In summary, the data seem to support our hypothesis that people with a good musical aptitude also will do well in speaker identification. This finding is consistent with those from other studies that have investigated this relationship (McGehee, 1944; Köster et al, 1997). However, since only a modest trend was discovered, it seems that there exists a need for further investigation in this area.

### Conclusion

Several relationships have been observed and a number of generalizations are possible. The first hypothesis -- that memory is positively correlated with the identification of speakers -- was not confirmed. Of course, whether or not an auditor attends to the stimulus defines the thoroughness of memory. This relationship -- between attention and memory -- also implies that in most criminal cases, the witness's memory of the event can be quite robust due to increased attention as a result of the threatening atmosphere. Whether the message itself was remembered or not, did not seem to be important. Verbal memory and other scores related to learning and recall of verbal material did not show a trend. The results for priming indicated that implicit or unconscious memory of a voice may contribute. However, it must be stressed that no statistically significant trends were found.

The second hypothesis -- that superior auditory skills correlate positively with earwitness identification -- was not supported with only the speech in noise test showing a small difference in the predicted direction. The parameters indicative of a person's basic

aural acuity, like frequency selectivity and temporal resolution, did not show any difference at all between the HIGH-SPID and LOW-SPID groups.

The obtained research results suggest that the first hypothesis is confirmed. That is, individuals who show a high degree of musical aptitude may be expected to do better in earwitness speaker identification than those who do not. Certain subparameters related to intonation seemed most important to this process.

It was found that a statistically significant speaker identification model can be constructed using the following parameters: attention/concentration, tonal memory, rhythm, priming, and timbre. Thus, it is concluded that, while these factors exhibited only modest trends and only one was actually significant, the small contributions made by several can result in a significant predictive model. The fact that the cited parameters contributed substantially to the identification model also indicates that they may be of assistance in evaluating a particular witness -- an issue of substantial recent interest to forensic phonetics. Although it was not the goal of this research to develop such a test, the findings of this study may be considered for this purpose or, at least as a basis for future research.

When the overall results are considered, two additional patterns emerge. First, it appears as if only those parameters which require high level (or more complicated processing) seem to show a positive correlation with earwitness speaker identification. With musical talent, for example, the tests that seem to be basic to the others and that require the least of cognitive processing by the listener, pitch and loudness, do not show a

positive trend. However, those that require more mental processing like tonal memory or the comparison of two rhythmic sequences, did show a trend. The same is true for the relationship between memory and tests which require substantial cognitive processing skills. Digit span backward, for example, seems to be a better predictor than digit span forward, the more basic version of the digit span. Finally, it is not surprising that attention/concentration also turned out to be a good addition to the speaker identification model. This process is crucial to tasks that require high level processing by listeners. In general, it seems as if speaker identification does not depend on simple processing attributes but rather on those that are more complicated and that require high level neural processing.

The second relationship that can be observed is that some of the parameters which appear to be good predictors of speaker identification accuracy, show right hemisphere dominance. Rhythm, tonal memory and timbre are all factors that have shown to be associated with activity in the right part of the brain (Wallin, 1991; Gordon, 1983). Note that the relationships cited above all assume right handed individuals: for them, it has been shown that language is processed mainly in the left hemisphere and, for example, music is associated with the right brain. So, if speaker identification shows a positive correlation with factors that are known to exhibit dominant right hemisphere activity, does that mean that also speaker identification is mainly an activity performed by the right side of the brain? Several investigators have studied or at least suggested this relationship (Campbell, 1992; Schacter and Church, 1992; Tartter, 1984; Van Lancker



and Kreiman, 1985; Young, 1983). Others have claimed that the left hemisphere operates on categorical or abstract auditory information (e.g., phonemes) and discards or ignores noncategorical information in the speech signal such as voice characteristics of a particular speaker. By contrast, the right hemisphere operates on noncategorical “acoustic gestalts” and to preserve information about prosodic features of speech, including characteristics of a particular speaker’s voice (Liberman, 1982; Mann and Liberman, 1983; Zaidel, 1985; Schacter and Church, 1992). Research on dichotic listening supports this suggestion (Blumstein and Cooper, 1974; Shipley-Brown et al, 1988). In both studies, it was found that complex acoustic stimuli such as intonation contours produced a reliable right hemisphere dominance. Further, findings from neuropsychological studies using patients with hemisphere lesions (Coslett et al, 1987; Ross, 1981), are consistent with the results from the dichotic studies.

Two research groups that directly investigated the link between hemisphere and speaker identification include 1) Van Lancker and Kreiman (1987, 1989) who studied patients with lesions in either the right or left hemisphere and Tartter (1994). Van Lancker and Kreiman found that patients with right-hemisphere lesions show deficits in voice recognition. In their dichotic listening study of 1988, they examined the voice recognition abilities of normal subjects and did not find an ear advantage. However, there was a relative left ear advantage for voice recognition compared with word recognition. Tartter (1984) who studied speaker identification using dichotic listening, demonstrated a non-significant right hemisphere advantage for speaker identification. Of course, the fact

that speaker identification can only be carried out by presenting verbal material, a type of stimulus that is associated with left hemisphere dominance, means that left side participation can not be excluded. However, Tartter showed right hemisphere participation that, although non-significant, exceeded the amount of activity measured on the left side. These results and earlier stated associations with processes that are right hemisphere dominated may suggest that speaker identification has an important right hemisphere component.

Only one parameter, that of tonal memory, was found to be statistically significant. This means, at this juncture anyway, that the only way to estimate the validity of an earwitness' judgment would be from scores of tonal memory. Since none of the other parameters were significant, it appears that the ability to identify speakers is a very complex process and one that is not easily assessed. However, certain of the other relationships and trends also suggest promising areas of research. They include intonation and musical aptitude; both of which appear useful to study in earwitness speaker identification in more detail. In addition, it was shown that certain skills in the field of memory might be useful when performing earwitness identification. Of course, only a limited number of parameters were studied here; there exist others which could very well be good predictors. Thus, the results of this investigation may serve as a guide for further study of appropriate relationships in the continuing search for predictors of earwitness speaker identification.

In summary, it appears that

- 1) Factors that require high level cognitive processing are better predictors of an earwitness' ability to identify speakers than those that are associated with basic mental skills and therefore,
- 2) Earwitnesses do not need to excel in the basic auditory and memory skills in order to carry out an identification task,
- 3) Earwitnesses that exhibit a high degree of musical aptitude can be expected to show better performance at identification of speakers than those that do not, and
- 4) Differences in intonation are important cues for identifying speakers for earwitnesses involved in a voice lineup.

## ABBREVIATIONS

|           |  |
|-----------|--|
| CAP       | Central auditory processing.   |
| F0        | Fundamental frequency.   |
| HIGH-SPID | The group consisting of the listeners with an identification score of 55% and above.           |
| LTS       | Long term spectrum.  |
| LOW-SPID  | The group consisting of the listeners with an identification score of 10% and below.           |
| PTA       | Pure Tone Average with PTA1 being the average of the thresholds for 500Hz, 1000Hz, and 2000Hz. |
| SD        | Standard deviation.  |
| SFF       | Speaking fundamental frequency.  |
| SPID      | Speaker identification.  |
| Z-score   | Difference from the mean expressed in standard deviation.                                      |

# APPENDIX A MEDICAL QUESTIONNAIRE

Hearing & Memory & SPID  
Gea DeJong

Subject: \_\_\_\_\_  
Date: \_\_\_\_\_

Your age: \_\_\_\_\_ Date of birth: \_\_\_\_\_  
Years of education, starting with high-school: \_\_\_\_\_  
Highest degree: \_\_\_\_\_  
Profession: \_\_\_\_\_

## QUESTIONNAIRE

### Medical history related to hearing:

|  | Yes   | No    |
|--|-------|-------|
| Have your ears ever been examined by an ear specialist or audiologist ?  | _____ | _____ |
| If so, when: _____   |       |       |
| What was the diagnosis ? _____   |       |       |
| Has either ear ever hurt or ached ?                                      | _____ | _____ |
| Which ear: _____ When _____  |       |       |
| Do you think your hearing has changed within the last six months ?       | _____ | _____ |
| Do you ever feel dizzy ?   | _____ | _____ |
| If yes, describe _____   |       |       |
| Do you have ringing in your ears ?                                       | _____ | _____ |
| If yes, describe _____   |       |       |
| Is one of your ears better than the other ?                              | _____ | _____ |
| If so, which is the better ?      Right _____      Left _____            |       |       |
| Which ear do you use the telephone ?      Right _____      Left _____    |       |       |
| Does your hearing seem better on some days than others ?                 | _____ | _____ |
| Do you experience frequent remarks from others concerning your hearing ? | _____ | _____ |
| Describe: _____  |       |       |
| Do you have trouble hearing on the phone ?                               | _____ | _____ |

|  | Yes   | No    |
|--|-------|-------|
| Do you have trouble hearing lectures ?   | _____ | _____ |
| Do you have trouble hearing in a group ?   | _____ | _____ |
| Do you have trouble hearing when talking to one person ?   | _____ | _____ |
| Have you ever been exposed to loud noises ?  | _____ | _____ |
| Describe: _____  |       |       |
| Do you have any difficulty in your present job/position because of your hearing ? If so, explain _____ | _____ | _____ |
| Is your working environment unusually noisy ?  | _____ | _____ |
| If yes, explain _____  |       |       |

#### Medical history related to memory:

Do you have a history of:

|  |       |       |
|--|-------|-------|
| head injury with loss of consciousness longer than 5 min.:                           | _____ | _____ |
| epilepsy:  | _____ | _____ |
| psychiatric problems (depression):   | _____ | _____ |
| and have you ever been hospitalized for it?  | _____ | _____ |
| diagnosed learning disability (reading, mathematics, etc.)                           | _____ | _____ |
| for which you received services like E.S.E.  |       |       |
| (Exceptional Student Education):   |       |       |
| drugs/alcohol that has caused physical/occupational problems:                        | _____ | _____ |
| any other neurological disease:  | _____ | _____ |
| If yes, describe _____   |       |       |
| Is there any significant other medical condition for which you are under treatment ? | _____ | _____ |
| If yes, describe _____   |       |       |

Medication:

---



---



---

**Musical/phonetics history**

Yes      No

Do you play an instrument ?

\_\_\_\_\_

Do you sing in a choir/band or have you been singing in a choir/band ?

\_\_\_\_\_

Is your study/job/hobby related to speech ?

\_\_\_\_\_

If yes explain: \_\_\_\_\_

Is your study/job/hobby related to phonetics ?

\_\_\_\_\_

If yes explain : \_\_\_\_\_

Is your study/job/hobby related to music ?

\_\_\_\_\_

If yes, explain: \_\_\_\_\_

APPENDIX B  
VOICE LINEUP SENTENCES

1. In half a day, he repaired five television sets, two telephones, and a very old stove.
2. Susie sewed zippers on two new dresses at Bessie's house.
3. Father asked how much money Tom had saved to buy a bird cage.
4. Ruth caught a cold because she wouldn't wear her new warm wool coat.
5. I found a huge toy music box outside Roy's house.

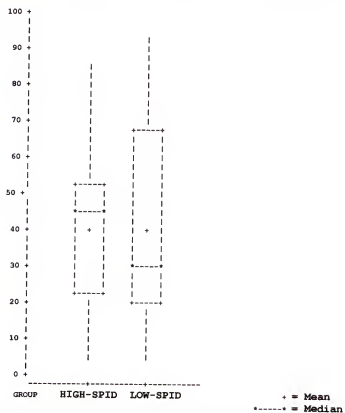


## APPENDIX C

### UNIVARIATE SAS PLOTS

## PITCH

|           |          |       |            |    |
|-----------|----------|-------|------------|----|
| HIGH-SPID | N        | 13    | 100% Max   | 84 |
|           | Mean     | 40.8  | 75% Q3     | 52 |
|           | Std Dev  | 23.2  | 50% Median | 45 |
|           | Skewness | 0.059 | 25% Q1     | 23 |
|           | Mode     | 45    | 0% Min     | 5  |
| LOW-SPID  | N        | 14    | 100% Max   | 92 |
|           | Mean     | 41.2  | 75% Q3     | 68 |
|           | Std Dev  | 29.9  | 50% Med    | 29 |
|           | Skewness | 0.707 | 25% Q1     | 20 |
|           | Mode     | 20    | 0% Min     | 6  |

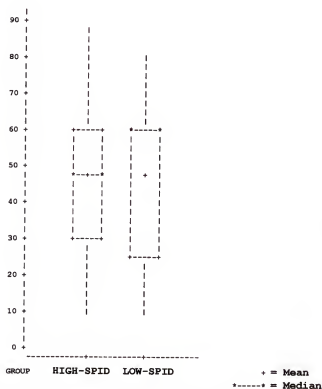


Descriptive univariate SAS-plot of the pitch data.

## LOUDNESS

|            |          |       |          |    |
|------------|----------|-------|----------|----|
| HIGH-SPID: | N        | 13    | 100% Max | 87 |
|            | Mean     | 47.1  | 75% Q3   | 59 |
|            | Std Dev  | 23.4  | 50% Med  | 48 |
|            | Skewness | 0.107 | 25% Q1   | 30 |
|            | Mode     | 30    | 0% Min   | 11 |

|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| LOW-SPID: | N        | 14     | 100% Max | 79 |
|           | Mean     | 46.3   | 75% Q3   | 59 |
|           | Std Dev  | 22.6   | 50% Med  | 59 |
|           | Skewness | -0.367 | 25% Q1   | 24 |
|           | Mode     | 59     | 0% Min   | 9  |

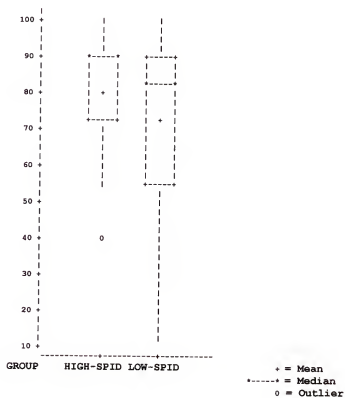


Descriptive univariate SAS-plot of the loudness data.

## RHYTHM

|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 99 |
|            | Mean     | 78.8   | 75% Q3   | 90 |
|            | Std Dev  | 19.5   | 50% Med  | 90 |
|            | Skewness | -0.814 | 25% Q1   | 73 |
|            | Mode     | 90     | 0% Min   | 39 |

|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| LOW-SPID: | N        | 14     | 100% Max | 99 |
|           | Mean     | 72.9   | 75% Q3   | 90 |
|           | Std Dev  | 27.1   | 50% Med  | 82 |
|           | Skewness | -1.070 | 25% Q1   | 55 |
|           | Mode     | 90     | 0% Min   | 12 |

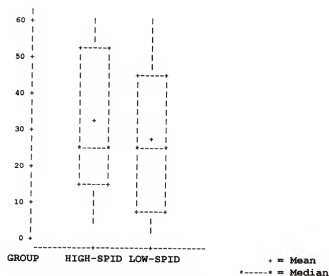


Descriptive univariate SAS-plot of the rhythm data.

## TIMBRE

|            |          |       |          |    |
|------------|----------|-------|----------|----|
| HIGH-SPID: | N        | 13    | 100% Max | 61 |
|            | Mean     | 32.9  | 75% Q3   | 52 |
|            | Std Dev  | 21.1  | 50% Med  | 25 |
|            | Skewness | 0.218 | 25% Q1   | 15 |
|            | Mode     | 61    | 0% Min   | 6  |

|           |          |       |          |    |
|-----------|----------|-------|----------|----|
| LOW-SPID: | N        | 14    | 100% Max | 61 |
|           | Mean     | 27.7  | 75% Q3   | 45 |
|           | Std Dev  | 21.4  | 50% Med  | 25 |
|           | Skewness | 0.335 | 25% Q1   | 8  |
|           | Mode     | 25    | 0% Min   | 2  |

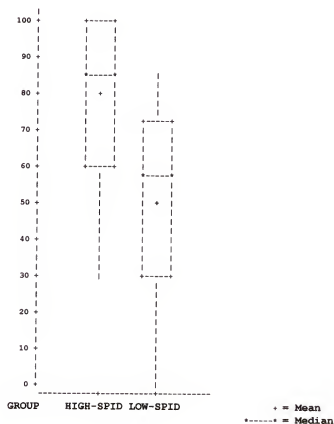


Descriptive univariate SAS-plot of the rhythm data.

## TONAL

|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 99 |
|            | Mean     | 78.8   | 75% Q3   | 99 |
|            | Std Dev  | 22.4   | 50% Med  | 85 |
|            | Skewness | -0.989 | 25% Q1   | 61 |
|            | Mode     | 99     | 0% Min   | 29 |

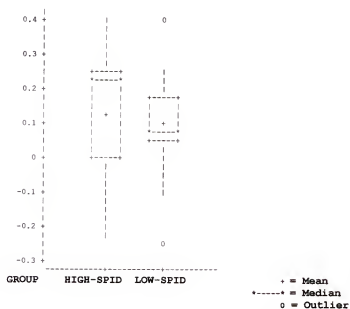
|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| LOW-SPID: | N        | 14     | 100% Max | 85 |
|           | Mean     | 50.0   | 75% Q3   | 72 |
|           | Std Dev  | 25.4   | 50% Med  | 57 |
|           | Skewness | -0.761 | 25% Q1   | 29 |
|           | Mode     | 52     | 0% Min   | 1  |



Descriptive univariate SAS-plot of the tonal memory data.

## PRIMING

|            |          |        |          |       |
|------------|----------|--------|----------|-------|
| HIGH-SPID: | N        | 13     | 100% Max | 0.4   |
|            | Mean     | 0.13   | 75% Q3   | 0.25  |
|            | Std Dev  | 0.177  | 50% Med  | 0.22  |
|            | Skewness | -0.573 | 25% Q1   | 0     |
|            | Mode     | 0.25   | 0% Min   | -0.23 |
| LOW-SPID:  | N        | 14     | 100% Max | 0.39  |
|            | Mean     | 0.09   | 75% Q3   | 0.18  |
|            | Std Dev  | 0.154  | 50% Med  | 0.08  |
|            | Skewness | -0.351 | 25% Q1   | 0.06  |
|            | Mode     | 0.06   | 0% Min   | -0.25 |

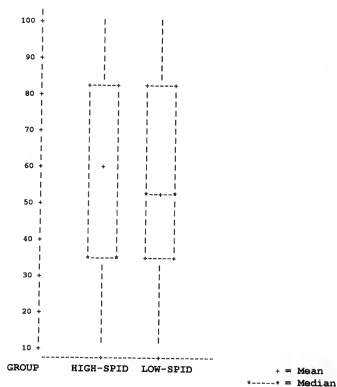


Descriptive univariate SAS-plot of the priming data.

## DIGIT SPAN FORWARD

|            |          |       |          |    |
|------------|----------|-------|----------|----|
| HIGH-SPID: | N        | 13    | 100% Max | 99 |
|            | Mean     | 58.8  | 75% Q3   | 82 |
|            | Std Dev  | 30.7  | 50% Med  | 36 |
|            | Skewness | 0.103 | 25% Q1   | 36 |
|            | Mode     | 36    | 0% Min   | 12 |

|           |          |       |          |    |
|-----------|----------|-------|----------|----|
| LOW-SPID: | N        | 14    | 100% Max | 99 |
|           | Mean     | 51.6  | 75% Q3   | 82 |
|           | Std Dev  | 28.6  | 50% Med  | 52 |
|           | Skewness | 0.382 | 25% Q1   | 36 |
|           | Mode     | 52    | 0% Min   | 12 |



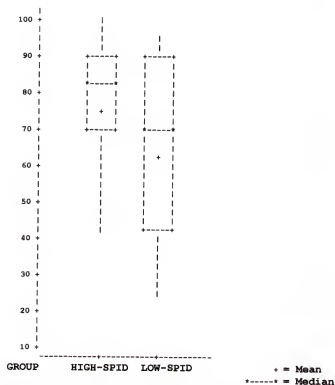
Descriptive univariate SAS-plot of the digit span forward data.



## DIGIT SPAN BACKWARD

|              |        |          |    |
|--------------|--------|----------|----|
| HIGH-SPID: N | 13     | 100% Max | 99 |
| Mean         | 75.2   | 75% Q3   | 90 |
| Std Dev      | 23.7   | 50% Med  | 82 |
| Skewness     | -1.724 | 25% Q1   | 70 |
| Mode         | 70     | 0% Min   | 14 |

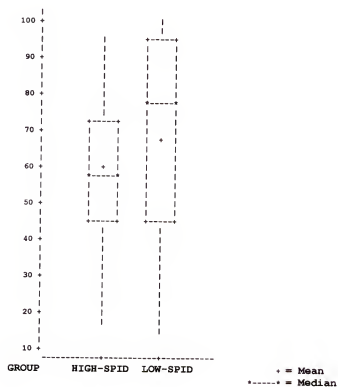
|             |        |          |    |
|-------------|--------|----------|----|
| LOW-SPID: N | 14     | 100% Max | 96 |
| Mean        | 63.4   | 75% Q3   | 90 |
| Std Dev     | 25.1   | 50% Med  | 70 |
| Skewness    | -0.434 | 25% Q1   | 42 |
| Mode        | 70     | 0% Min   | 26 |



Descriptive univariate SAS-plot of the digit span backward data.

## LOGICAL MEMORY I

|            |          |        |          |      |
|------------|----------|--------|----------|------|
| HIGH-SPID: | N        | 13     | 100% Max | 96   |
|            | Mean     | 59.7   | 75% Q3   | 73   |
|            | Std Dev  | 25.1   | 50% Med  | 57   |
|            | Skewness | -0.171 | 25% Q1   | 46   |
|            | Mode     | 57     | 0% Min   | 17   |
| LOW-SPID:  | N        | 14     | 100% Max | 99   |
|            | Mean     | 68.7   | 75% Q3   | 96   |
|            | Std Dev  | 30.8   | 50% Med  | 78.5 |
|            | Skewness | -0.601 | 25% Q1   | 46   |
|            | Mode     | 96     | 0% Min   | 16   |

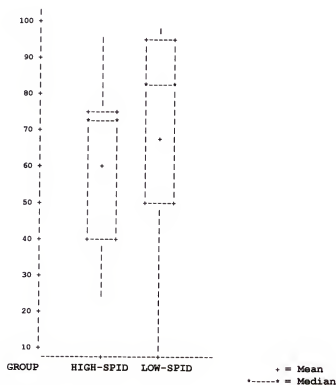


Descriptive univariate SAS-plot of the logical memory I data.

## LOGICAL MEMORY II

|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 94 |
|            | Mean     | 61.2   | 75% Q3   | 75 |
|            | Std Dev  | 23.6   | 50% Med  | 72 |
|            | Skewness | -0.202 | 25% Q1   | 40 |
|            | Mode     | 75     | 0% Min   | 24 |

|           |          |        |          |    |
|-----------|----------|--------|----------|----|
| LOW-SPID: | N        | 14     | 100% Max | 97 |
|           | Mean     | 67     | 75% Q3   | 94 |
|           | Std Dev  | 31.6   | 50% Med  | 82 |
|           | Skewness | -0.710 | 25% Q1   | 49 |
|           | Mode     | 53     | 0% Min   | 10 |

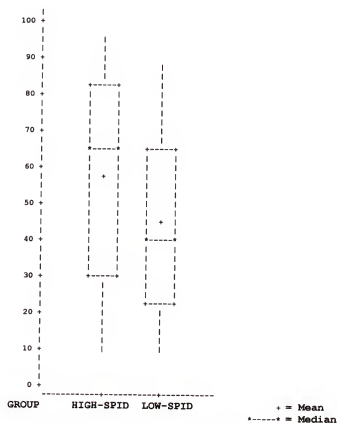


Descriptive univariate SAS-plot of the logical memory II data.

## ATTENTION/CONCENTRATION

|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 94 |
|            | Mean     | 56.4   | 75% Q3   | 82 |
|            | Std Dev  | 28.0   | 50% Med  | 65 |
|            | Skewness | -0.343 | 25% Q1   | 31 |
|            | Mode     | 82     | 0% Min   | 9  |

|           |          |       |          |    |
|-----------|----------|-------|----------|----|
| LOW-SPID: | N        | 14    | 100% Max | 87 |
|           | Mean     | 43.9  | 75% Q3   | 64 |
|           | Std Dev  | 24.0  | 50% Med  | 41 |
|           | Skewness | 0.197 | 25% Q1   | 22 |
|           | Mode     | 15    | 0% Min   | 9  |

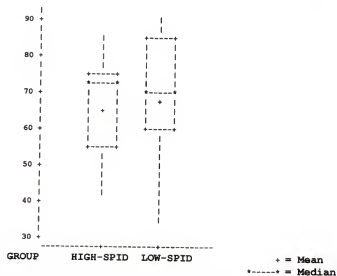


Descriptive univariate SAS-plot of the attention/concentration data.

## DELAYED RECALL

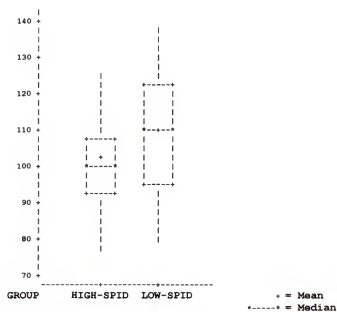
|            |          |        |          |    |
|------------|----------|--------|----------|----|
| HIGH-SPID: | N        | 13     | 100% Max | 86 |
|            | Mean     | 66.2   | 75% Q3   | 74 |
|            | Std Dev  | 13.7   | 50% Med  | 72 |
|            | Skewness | -0.227 | 25% Q1   | 54 |
|            | Mode     | 74     | 0% Min   | 43 |

|           |          |         |          |      |
|-----------|----------|---------|----------|------|
| LOW-SPID: | N        | 14      | 100% Max | 91   |
|           | Mean     | 68.6    | 75% Q3   | 85   |
|           | Std Dev  | 19.0    | 50% Med  | 70.5 |
|           | Skewness | -0.5092 | 25% Q1   | 59   |
|           | Mode     | 62      | 0% Min   | 34   |



## VERBAL MEMORY

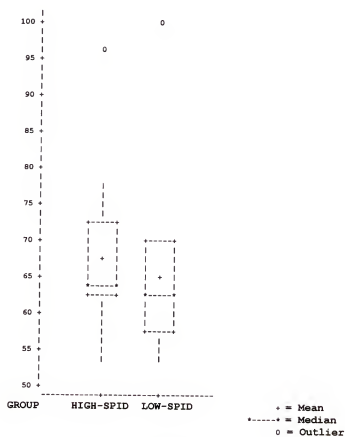
|            |          |        |          |     |
|------------|----------|--------|----------|-----|
| HIGH-SPID: | N        | 13     | 100% Max | 124 |
|            | Mean     | 101.4  | 75% Q3   | 108 |
|            | Std Dev  | 14.2   | 50% Med  | 101 |
|            | Skewness | -0.084 | 25% Q1   | 93  |
|            | Mode     | 101    | 0% Min   | 78  |
| LOW-SPID:  | N        | 14     | 100% Max | 138 |
|            | Mean     | 109.1  | 75% Q3   | 123 |
|            | Std Dev  | 17.3   | 50% Med  | 110 |
|            | Skewness | -0.118 | 25% Q1   | 96  |
|            | Mode     | 123    | 0% Min   | 80  |



Descriptive univariate SAS-plot of the delayed recall data.

## MRT

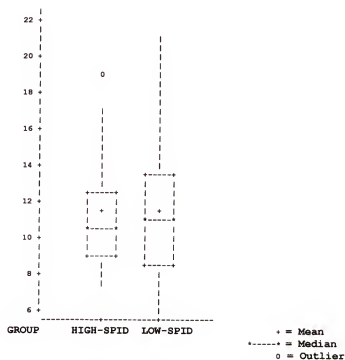
|            |          |       |          |     |
|------------|----------|-------|----------|-----|
| HIGH-SPID: | N        | 13    | 100% Max | 96  |
|            | Mean     | 68    | 75% Q3   | 72  |
|            | Std Dev  | 10.7  | 50% Med  | 64  |
|            | Skewness | 1.530 | 25% Q1   | 62  |
|            | Mode     | 60    | 0% Min   | 54  |
| LOW-SPID:  | N        | 14    | 100% Max | 100 |
|            | Mean     | 65    | 75% Q3   | 70  |
|            | Std Dev  | 11.6  | 50% Med  | 63  |
|            | Skewness | 2.084 | 25% Q1   | 58  |
|            | Mode     | 70    | 0% Min   | 54  |



Descriptive univariate SAS-plot of the MRT data.

## GAP DETECTION

|            |          |       |          |       |
|------------|----------|-------|----------|-------|
| HIGH-SPID: | N        | 13    | 100% Max | 19.20 |
|            | Mean     | 11.35 | 75% Q3   | 12.27 |
|            | Std Dev  | 3.49  | 50% Med  | 10.32 |
|            | Skewness | 1.273 | 25% Q1   | 8.83  |
|            | Mode     | 7.722 | 0% Min   | 7.72  |
| LOW-SPID:  | N        | 14    | 100% Max | 20.91 |
|            | Mean     | 11.34 | 75% Q3   | 13.58 |
|            | Std Dev  | 3.89  | 50% Med  | 10.80 |
|            | Skewness | 1.002 | 25% Q1   | 8.68  |
|            | Mode     | 6.042 | 0% Min   | 6.04  |



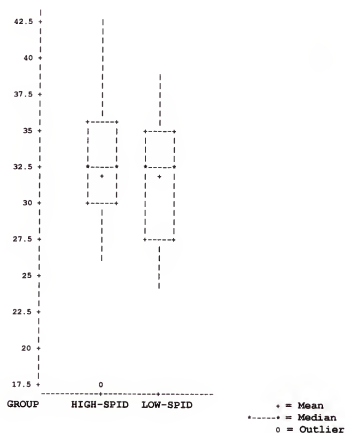
Descriptive univariate SAS-plot of the gap detection data.



## FREQUENCY SELECTIVITY

|            |          |        |          |      |
|------------|----------|--------|----------|------|
| HIGH-SPID: | N        | 13     | 100% Max | 42.3 |
|            | Mean     | 31.9   | 75% Q3   | 35.5 |
|            | Std Dev  | 5.98   | 50% Med  | 32.4 |
|            | Skewness | -0.867 | 25% Q1   | 30.1 |
|            | Mode     | 17.46  | 0% Min   | 17.5 |

|           |          |        |          |      |
|-----------|----------|--------|----------|------|
| LOW-SPID: | N        | 14     | 100% Max | 38.7 |
|           | Mean     | 32.0   | 75% Q3   | 34.8 |
|           | Std Dev  | 4.69   | 50% Med  | 32.7 |
|           | Skewness | -0.217 | 25% Q1   | 27.6 |
|           | Mode     | 38.69  | 0% Min   | 24.6 |



Descriptive univariate SAS-plot of the frequency selectivity data.

# APPENDIX D CORRELATION ANALYSIS

Correlation Analysis: HIGH-SPID  
Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 13

|         | RHYTHM             | TIMBRE             | TONAL              | PRIMING            |
|---------|--------------------|--------------------|--------------------|--------------------|
| RHYTHM  | 1.00000<br>0.0     | -0.05037<br>0.8702 | 0.04253<br>0.8903  | 0.17735<br>0.5622  |
| TIMBRE  | -0.05037<br>0.8702 | 1.00000<br>0.0     | 0.30613<br>0.3090  | -0.29060<br>0.3354 |
| TONAL   | 0.04253<br>0.8903  | 0.30613<br>0.3090  | 1.00000<br>0.0     | -0.42584<br>0.1468 |
| PRIMING | 0.17735<br>0.5622  | -0.29060<br>0.3354 | -0.42584<br>0.1468 | 1.00000<br>0.0     |
| DIGFOR  | -0.14402<br>0.6388 | -0.03039<br>0.9215 | 0.20166<br>0.5088  | -0.02787<br>0.9280 |
| DIGBACK | -0.06037<br>0.8447 | -0.13314<br>0.6646 | 0.07447<br>0.8090  | 0.23874<br>0.4321  |
| ATTENT  | -0.03755<br>0.9031 | -0.16322<br>0.5942 | -0.09186<br>0.7654 | 0.35002<br>0.2410  |
| MRT     | 0.00160<br>0.9958  | 0.08435<br>0.7841  | 0.05430<br>0.8602  | -0.23074<br>0.4482 |
|         | DIGFOR             | DIGBACK            | ATTENT             | MRT                |
| RHYTHM  | -0.14402<br>0.6388 | -0.06037<br>0.8447 | -0.03755<br>0.9031 | 0.00160<br>0.9958  |
| TIMBRE  | -0.03039<br>0.9215 | -0.13314<br>0.6646 | -0.16322<br>0.5942 | 0.08435<br>0.7841  |
| TONAL   | 0.20166<br>0.5088  | 0.07447<br>0.8090  | -0.09186<br>0.7654 | 0.05430<br>0.8602  |
| PRIMING | -0.02787<br>0.9280 | 0.23874<br>0.4321  | 0.35002<br>0.2410  | -0.23074<br>0.4482 |
| DIGFOR  | 1.00000<br>0.0     | 0.61200<br>0.0262  | 0.75112<br>0.0031  | 0.07272<br>0.8134  |
| DIGBACK | 0.61200<br>0.0262  | 1.00000<br>0.0     | 0.83918<br>0.0003  | -0.64297<br>0.0178 |
| ATTENT  | 0.75112<br>0.0031  | 0.83918<br>0.0003  | 1.00000<br>0.0     | -0.20429<br>0.5032 |
| MRT     | 0.07272<br>0.8134  | -0.64297<br>0.0178 | -0.20429<br>0.5032 | 1.00000<br>0.0     |

Correlation Analysis LOW-SPID  
Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 14

|         | RHYTHM             | TIMBRE             | TONAL              | PRIMING            |
|---------|--------------------|--------------------|--------------------|--------------------|
| RHYTHM  | 1.00000<br>0.0     | 0.05532<br>0.8510  | -0.11822<br>0.6873 | 0.00910<br>0.9754  |
| TIMBRE  | 0.05532<br>0.8510  | 1.00000<br>0.0     | 0.42277<br>0.1321  | -0.16051<br>0.5836 |
| TONAL   | -0.11822<br>0.6873 | 0.42277<br>0.1321  | 1.00000<br>0.0     | -0.24111<br>0.4063 |
| PRIMING | 0.00910<br>0.9754  | -0.16051<br>0.5836 | -0.24111<br>0.4063 | 1.00000<br>0.0     |
| DIGFOR  | 0.37354<br>0.1883  | 0.57733<br>0.0306  | 0.27417<br>0.3428  | -0.40019<br>0.1562 |
| DIGBACK | 0.11760<br>0.6889  | 0.22872<br>0.4316  | 0.07720<br>0.7931  | 0.48377<br>0.0797  |
| ATTENT  | 0.31581<br>0.2714  | 0.45840<br>0.0993  | 0.00101<br>0.9973  | 0.11718<br>0.6899  |
| MRT     | -0.09376<br>0.7499 | -0.10483<br>0.7213 | -0.35037<br>0.2194 | 0.38497<br>0.1741  |
|         | DIGFOR             | DIGBACK            | ATTENT             | MRT                |
| RHYTHM  | 0.37354<br>0.1883  | 0.11760<br>0.6889  | 0.31581<br>0.2714  | -0.09376<br>0.7499 |
| TIMBRE  | 0.57733<br>0.0306  | 0.22872<br>0.4316  | 0.45840<br>0.0993  | -0.10483<br>0.7213 |
| TONAL   | 0.27417<br>0.3428  | 0.07720<br>0.7931  | 0.00101<br>0.9973  | -0.35037<br>0.2194 |
| PRIMING | -0.40019<br>0.1562 | 0.48377<br>0.0797  | 0.11718<br>0.6899  | 0.38497<br>0.1741  |
| DIGFOR  | 1.00000<br>0.0     | 0.34903<br>0.2213  | 0.78352<br>0.0009  | -0.03116<br>0.9158 |
| DIGBACK | 0.34903<br>0.2213  | 1.00000<br>0.0     | 0.61808<br>0.0185  | 0.10407<br>0.7233  |
| ATTENT  | 0.78352<br>0.0009  | 0.61808<br>0.0185  | 1.00000<br>0.0     | 0.30312<br>0.2921  |
| MRT     | -0.03116<br>0.9158 | 0.10407<br>0.7233  | 0.30312<br>0.2921  | 1.00000<br>0.0     |

APPENDIX E  
INTERACTION ANALYSIS

Logistic Linear Regression  
Analysis of Maximum Likelihood Estimates

| Combination           | Parameter<br>Estimate | p-values |
|-----------------------|-----------------------|----------|
| Rhythm*Timbre         | 0.0858                | 0.3641   |
| Rhythm*Tonal Mem      | 0.0122                | 0.7939   |
| Rhythm*Priming        | 6.4643                | 0.5872   |
| Rhythm*DigitSpanB     | -0.0839               | 0.3997   |
| Rhythm*Attent/Con     | -0.0165               | 0.7679   |
| Rhythm*MRT            | -0.1068               | 0.4139   |
| Timbre*Tonal Mem      | 0.0005                | 0.8456   |
| Timbre*Priming        | -0.0137               | 0.9861   |
| Timbre*DigitSpanB     | 0.0057                | 0.4884   |
| Timbre*Attent/Con     | -0.0070               | 0.2553   |
| Timbre*MRT            | 0.0074                | 0.5511   |
| Tonal Mem*Priming     | -6.9683               | 0.2424   |
| Tonal Mem*DigitSpanB  | 0.0239                | 0.4794   |
| Tonal Mem*Attent/Con  | -0.0776               | 0.1562   |
| Tonal Mem*MRT         | 0.1143                | 0.1786   |
| Priming*DigitSpanB    | -4.7490               | 0.6710   |
| Priming*Attent/Con    | 4.7877                | 0.7339   |
| Priming*MRT           | -4.0878               | 0.4758   |
| DigitSpanB*Attent/Con | 0.0301                | 0.2465   |
| DigitSpanB*MRT        | -0.2372               | 0.2251   |
| Attent/Con*MRT        | -0.00311              | 0.3247   |

APPENDIX F  
ESTIMATES AND P-VALUES OF THE SECOND MODEL

Logistic Linear Regression  
Analysis of Maximum Likelihood Estimates

---

---

| Combination             | Parameter<br>Estimate | p-values |
|-------------------------|-----------------------|----------|
| Rhythm                  | 0.1881                | 0.1720   |
| Timbre                  | -0.2240               | 0.2073   |
| Tonal Mem               | 1.7004                | 0.1279   |
| Priming                 | 39.6338               | 0.1545   |
| DigitSpanB              | -0.1025               | 0.2736   |
| Attention/Conc          | 1.5871                | 0.1302   |
| Tonal Mem * Attent/Conc | -0.0193               | 0.1248   |

---

| Criterion | Intercept<br>Intercept and |            | Chi-Square for Covariates   |
|-----------|----------------------------|------------|-----------------------------|
|           | Only                       | Covariates |                             |
| AIC       | 39.393                     | 27.887     | .                           |
| SC        | 40.689                     | 38.254     | .                           |
| -2 LOG L  | 37.393                     | 11.887     | 25.505 with 7 DF (p=0.0006) |
| Score     | .                          | .          | 11.407 with 7 DF (p=0.1218) |

---

---

Note: the p-value is the p-value for this entire model

## LIST OF REFERENCES

- Abberton, E. (1974) Listener identification of speakers from larynx frequency, *Work in Progress*, Department of Phonetics and Linguistics, University College, London.
- Abberton, E. and Fourcin, A. (1978) Intonation and speaker identification, *Language and Speech*, 21:305-318.
- Agresti, A. (1984) *Analysis of Ordinal Categorical Data*, New York: John Wiley & Sons, Inc.
- Alain, C. and Woods, D.L. (1997) Attention modulates auditory pattern memory as indexed by event-related brain potentials, *Psychophysiology*, 34(5): 534-546.
- American National Standards Institute, Methods for Manual Pure-Tone Threshold Audiometry, *ANSI S3.21-1978* (R 1991) New York.
- American Speech-Language-Hearing Association (ASHA), (1990) *Guidelines for screening for hearing impairment and middle-ear disorders*. American Speech and Hearing Association, 32: 17-24.
- Atal, B.S. (1972) Automatic speaker recognition based on pitch contours, *Journal of the Acoustical Society of America*, 52:1687-1697.
- Atal, B.S. (1974) Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, *Journal of the Acoustical Society of America*, 55:1304-1312.
- Atkinson, R.C. and Shiffrin, R.M (1965) Mathematical models for memory and learning. *Technical Report 79*, Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Atkinson, R. C. and Shiffrin, R.M. (1968) Human memory: A proposed system and its control processes, *Advances in the psychology of learning and motivation*, K.W. Spence and J.T. Spence (Eds.) Vol.2 New York: Academic Press.
- Atkinson, R.C. and Shiffrin, R.M (1971) The control of short-term memory, *Scientific American*, August, 82-90.

- Atwood, W. and Hollien, H. (1986) Stress monitoring by polygraph for research purposes. *Polygraph*, 15: 47-56.
- Babcock, H. and Levy, L (1940) *Test and Manual of Directions: The Revised Examination for the Measurement of Efficiency of Mental Functioning*. Chicago: Stoelting.
- Bannatyne, A.D., and Wichiarajote, P. (1969) Relationship between written spelling, motor functioning and sequencing skills, *Journal of Learning Disabilities*, 2:6-18.
- Barclay, C.R. (1986) Schematization of autobiographical memory, in D.C. Rubin (Ed.), *Autobiographical Memory*, 82-99, Cambridge, England: Cambridge University Press.
- Bartholomeus, B. (1973) Voice identification by nursery school children, *Canadian Psychology*, 27:464-72.
- Bartlett, F.C. (1932) *Remembering*, Cambridge, England: Cambridge University Press.
- Belli, R.F. (1989) Influences of misleading post-event information: Misinformation interference and acceptance, *Journal of Experimental Psychology: General*, 118:72-85.
- Bennani, Y. and Gallinari, P. (1995) Neural networks for discrimination and modelization of speakers, *Speech Communication*, 17: 159-175.
- Blumstein, S. and Cooper, W.E. (1974) Hemispheric processing of intonation contours, *Cortex*, 10: 146-158.
- Boder, E. (1973) Developmental dyslexia: a diagnostic approach based on three atypical reading-spelling patterns, *Devision of Medical Child Neurology*, 15:663-687.
- deBoer, E., and Bouwmeester, J. (1974) Critical bands and sensorineural hearing loss, *Audiology*, 13:236-259.
- Bolt, R.H., Cooper, F.S., David, E.E., Denes, P.B., Pickett, J.M. and Stevens, K.N., (1970) Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes, *Journal of the Acoustical Society of America*, 47:597-612.
- Bolt, R.H., Cooper, F.S., David, E.C., Denes, P.B., Pickett, J.M. and Stevens, K.N. (1973) Speaker identification by speech spectrograms: some further observations, *Journal of the Acoustical Society of America*, 54:531-534.

- Bonding, P. (1979) Frequency selectivity and speech discrimination in sensorineural hearing loss, *Scandinavian Audiology*, 8:205-216.
- Botwinick, J. and Storandt, M. (1974) *Memory, Related Functions and Age*, Springfield, IL: Charles C. Thomas.
- Brewer, W.F (1986) What is autobiographical memory ? In D.C. Rubin (Ed.), *Autobiographical Memory*, 25-49, Cambridge, England: Cambridge University Press.
- Brewin, C. R. (1997) Recovered memories, *British Journal of Psychiatry*, 170(4): 388.
- Bricker, P.D., Gnanadesikan, R., Mathews, M.V., Pruzansky, S., Tukey, P.A., Wachter, K.W. and Warner, J.L. (1971) Statistical techniques for talker identification, *Bell System Technology Journal*, 50:1427-1450.
- Bricker, P.D. and Pruzansky, S. (1966) Effects of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America*, 40:1441-1448.
- Broeders, A.P.A. (1996) Earwitness identification: common ground, disputed territory and uncharted areas, *Forensic Linguistics*, 3(1):1-13.
- Broeders, A.P.A. and Rietveld, A.C.M. (1995) Speaker identification by earwitnesses, *BEIPHOL, Studies in Forensic Phonetics*, 64:24-40.
- Bromley, D.B. (1958) Some effects of age on short-term learning and memory, *Journal of Gerontology*, 13:398-406.
- Brown, R. (1979) Memory and decision in speaker recognition, *International Journal for Man-machine Studie*, 11:729-942.
- Bull, R. and Clifford, B. (1984) Earwitness voice recognition accuracy, *Eyewitness Testimony* (Wells and Loftus, Eds.), Cambridge: Cambridge University Press, 92-123.
- Bull, R., Rathborn, H., and Clifford, B. (1983) The voice recognition accuracy of blind listeners, *Perception*, 12: 223-6.
- Butler, S.R. and Glass, A. (1974) Asymmetries in the electroencephalogram associated with cerebral dominance. *EEG and Clinical Neurophysiology*, 36: 481-491.
- Caffee, J.M. (1997) *Measures of Distortion Product Otoacoustic Emissions and Frequency Resolution in Clinically Normal Hearing Subjects With a History of Noise Exposure*, Master's Thesis, University of Florida.



- Calinski, T., Jassem, W. And Kaczmarck, Z. (1970) Investigation of vowel formant frequencies as personal voice characteristics by means of multivariate analysis of variance, *Speech Analysis and Synthesis* (W. Jassem, Ed.), Warsaw, Poland, 2:7-40.
- Campbell, D.G. (1992) *Introduction to the Musical Brain*, St. Louis, MO: MMB Music Inc.
- Carbonell, J.R., Grignetti, M.C., Stevens, K.N., Williams, C.E. and Woods, B. (1965) Speaker authentication techniques, *Report 1296 No. DA-28-043-AMC-00116(E)*, Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts.
- Carey, M. J., Parris, E.S., Bennett, S.J., Lloyd-Thomas, H. (1997) Comparison of model estimation techniques for speaker verification, *Speech Processing ICASSP*, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2) 97CB36052: 1083-1086.
- Carmon, A., Lavy, S., Gordon, H. and Portnoy, Z. (1975) Hemispheric differences in rCBF during verbal and non-verbal tasks. *Brain Work* (D.H. Ingvar and N.A. Lassen, eds.), Alfred Benzon Symposium VIII, Copenhagen: 414-423.
- Carterette, E. and Barnebey, A. (1975) Recognition memory for voices, In A. Cohen and G. Nooteboom (Eds.), *Structure and processes in speech perception*, New York: Springer.
- Casella, G. and Berger, R.L. (1990) *Statistical Inference*, Pacific Grove, California: Wadsworth & Brooks.
- Castellano, P., Slomka, S., Barger, P. (1997) Gender gates for telephone-based automatic speaker recognition. *Digital Signal Processing: A Review Journal* (7)2: 65-79.
- Christianson, S. and Engelberg, E. (1997) Remembering and forgetting traumatic experiences: A matter of survival. , *Recovered Memories and False Memories. Debates in Psychology*. (M.A. Conway, Ed.), Oxford, England: Oxford University Press, 230-250.
- Church, B.A. and Schacter, D.L (1994) Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20: 521-533.
- Clarke, F.R. and Becker, R.W. (1969) Comparison of techniques for discriminating among talkers, *Journal of Speech and Hearing Research*, 12:747-761.

- Clifford, B.R. (1980) Voice identification by human listeners: on earwitness reliability, *Law Human Behavior*, 4:373-394.
- Clifford, B.R., and Denot, H. (1982) *Visual and verbal testimony and identification under conditions of stress*, Unpublished manuscript, North East London Polytechnic.
- Clifford, B.R., Rathborn, H. and Bull, H. (1981) The effects of delay on voice recognition accuracy, *Law Human Behavior*, 5:201-208.
- Cole, R.A., Coltheart, M. and Allard, F. (1974) Memory of a speaker's voice: Reaction time to same- or different-voiced letters, *Quarterly Journal of Experimental Psychology*, 26(1):1-7.
- Coleman, R. (1973) Speaker identification in the absence of inter-subject differences in glottal source characteristics, *Journal of the Acoustical Society of America*, 53: 1741-3.
- Coleman, R. and Lass, N. (1981) Effect of prior exposure to stimulus material on identification of speakers' sex, height and weight, *Perceptual and Motor Skills*, 52: 619-622.
- Colwell, R. (1963) An investigation of musical achievement among vocal students, vocal-instrumental students and instrumental students. *Journal of Research in Music Education*, 11:123-130.
- Compton, A.J. (1963) Effects of filtering and vocal duration upon the identification of speakers, aurally, *The Journal of the Acoustical Society of America*, 35(11):1748-1952.
- Constanzo, F.S., Markel, N.N. and Costanzo, P.R. (1969) Voice quality profile and perceived emotion, *Journal of Counsel Psychology*, 16:267-270.
- Conway, M.A. (1990) *Autobiographical Memory: An Introduction*, Philadelphia: Open University Press.
- Conway, M.A. (1995) *Recovered Memories and False Memories*. Rutgers Symposium on Self & Social Identity, 1995, Oxford, England: Oxford University Press.
- Cort, S. and Murry, T. (1972) Aural identification of children's voices, *Journal of the Acoustical Society of America*, 51:131.
- Coslett, H.B., Roeltgen, D.P., Rothi, L.G. and Heilman, K.M (1987) Transcortical sensory aphasia: Evidence for subtypes, *Brain and Language*, 32: 362-378.

- Craik, F.I.M. (1977) Age differences in human memory. In J.E. Birren and K.W. Schaie (Eds.), *Handbook of the Psychology of Aging*, 384-420, New York: Van Nostrand Reinhold.
- Craik, F.I.M. and McDowd, J.M. (1987) Age differences in recall and recognition, *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13:474-479.
- Darwin, C.J. and Bethell-Fox, C.E. (1977) Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 3: 665-672.
- Deffenbacher, K., Brown, E.L., and Sturgill, W. (1978) Some predictors of eyewitness memory accuracy, In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (Eds.), *Practical aspects of memory*, London: Academic Press.
- DeJong, G. (1996) *Speaker Identification Accuracy and Earwitness Characteristics*, presented at the International Association of Forensic Phonetics Conference, Wiesbaden, Germany, July.
- Dirks, D. Morgan, D., and Dubno, J. (1982) A procedure for quantifying the effects of noise on speech recognition, *Journal of Speech and Hearing Disorders*, 47:114-123.
- Doddington, G.R. (1970) A method of speaker verification, Ph.D. Dissertation, University of Wisconsin.
- Doddington, G.R. (1971) A method of speaker verification, *Journal of the Acoustical Society of America*, 49:139(A)
- Doddington, G.R. (1974) Speaker verification, *Final Report RADC-TR-74-179*, Griffis Air Force Base, Rome, NY.
- Doddington, G.R. (1976) Personal identity verification using voice, *Proceedings of ELECTRO-76*, 22-4:1-5.
- Doherty, E.T. (1976) An evaluation of selected acoustic parameters for use in speaker identification, *Journal of Phonetics*, 4:321-326.
- Doherty, E.T. and Hollien, H. (1978) Multiple factor speaker identification of normal and distorted speech, *Journal of Phonetics*, 6:1-8.
- Drake, R.M. (1931) *Test of musical talent*, Doctoral dissertation, University of London.

- Dreschler, W.A. and Plomp, R. (1980) Relations between psychophysical data and speech perception for hearing impaired subjects, *Journal of the Acoustical Society of America*, 68:1608-1615.
- Duquesnoy, A. (1983) The intelligibility of sentences in quiet and in noise in aged listeners, *Journal of the Acoustical Society of America*, 74:1136-1144.
- Ebbinghaus, H. (1885) *Über das Gedächtnis: Untersuchungen zur Experimentellen Psychologie*. Leipzig: Dunker and Humbolt. Translated by H.A. Ruger and C.E. Byssensine as *Memory: A contribution to experimental psychology*, New York: Dover, 1913.
- Elkins, E.F. (1971) Evaluation of modified rhyme test results from impaired- and normal-hearing listeners, *Journal of Speech and Hearing Research*, 14(3): 589-595.
- Elliott, L. (1963) Prediction of speech discrimination scores from other test information, *Journal of Auditory Research*, 3: 35-45.
- Endress, W., Bambach, W. and Flosser, G. (1971) Voice spectrograms as a function of age, voice disguise and voice imitation, *Journal of the Acoustical Society of America*, 49:1842-1848.
- Erickson, R.C. and Scott, M.L. (1977) Clinical memory testing: A review, *Psychological Bulletin*, 84(6): 1130-1149.
- Evans, E.F. (1978) Peripheral auditory processing in normal and abnormal ear: physiological considerations for attempts to compensate for auditory deficits by acoustic and electrical prostheses, *Scandinavian Audiology*, Supplementum, 6: 9-47.
- Fairbanks, G. and Hoaglin, L.W. (1941) An experimental study of the durational characteristics of the voice during the expression of emotion, *Speech Monographs*, 8:85-90.
- Fairbanks, G. and Pronovost, W. (1939) An experimental study of the pitch characteristics of the voice during the expression of emotion, *Speech Monographs*, 6:87-104.
- Fant, G.M. (1956) On the predictability of formant levels and spectrum envelopes from formant frequencies. For Roman Jakobson (eds. M. Halle, H. Lunt and H. Maclean), The Hague: Mouton
- Farnsworth, P.R. (1938) Auditory acuity and musical ability in the first four grades. *Journal of Psychology*, 6: 95-98.

- Farnsworth, P.R. (1941) Further data on the Adlerian theory of artistry, *Journal of General Psychology*, 24: 447-450.
- Fieldhouse, A.E. (1937) *A Study of Backwardness in Singing Among School Children*, Unpublished doctoral dissertation, London University, 1937.
- deFigueiredo, R.M. and deSouza Britto, H. (1996) A report on the acoustic effects of one type of disguise, *Forensic Linguistics, The International Journal of Speech, Language and the Law*, 3(1):168-175.
- Francis, L.J. (1992) The influence of religion, gender, and social class on attitudes toward school among 11-year-olds in England. *Journal of Experimental Education*, 60(4): 339-348.
- Fried, I. Tanguay, P.E., Boder, E. Doubleday, C. and Greensite, M. (1981) Developmental dyslexia: electrophysiological evidence of clinical subgroups, *Brain and Language*, 12:14-22.
- Friedhoff, A.J., Alpert, M. and Kurtzberg, R.L. (1964) An electroacoustical analysis of the effects of stress on voice, *Journal of Neuropsychiatry* 5:265-272.
- Friedlander, B. (1970) Receptive language development: Issues and problems, *Merrill-Palmer Quarterly of Behavior and Development*, 16:7-15.
- Furui, S. (1978) Effects of long-term spectral variability on speaker recognition, *Journal of the Acoustical Society of America*, 64:S183.
- Furui, S. (1981) Cepstral analysis technique for automatic speaker verification, *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-29:254-272.
- Galin, D. and Ornstein, R. (1972) Lateral specialization of cognitive mode: an EEG study. *Psychophysiology*, 9: 412-418.
- Gass, C.S. and Apple, C. (1997) Cognitive complaints in closed-head injury: Relationship to memory test performance and emotional disturbance, *Journal of Clinical & Experimental Neuropsychology*, 19(2) 290-299.
- Gernsbacher, M.A. (1985) Surface information loss in comprehension, *Cognitive Psychology*, 17:324-363.
- Gfroerer, S. (1994) Häufigkeit und Art forensischer Stimmverstellungen, Unpublished paper.

- Gibson, J.J. and Gibson, E. (1955) Perceptual learning: Differentiation or enrichment?, *Psychological Review*, 62:32-41.
- Glasberg, B.R. and Moore, B.C.J. (1989) Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech, *Scandinavian Audiology Supplementum* 32:1-25.
- Goggin, J.P., Thompson, C.P. Strube, G. and Simental, L.R. (1991) The role of language familiarity in voice identification, *Memory and Cognition*, 19:448-458.
- Goldinger, S.D. (1996) Words and voices: Episodic traces in spoken word identification and recognition memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5):1166-1183.
- Gordon, E. (1965) *Musical Aptitude Profile*, Boston: Houghton Mifflin.
- Gordon, H.W. (1983) Music and the right hemisphere. *Functions of the Right Cerebral Hemisphere* (A.W. Young, ed.), London: Academic Press, 65-86.
- Graf, P. and Williams, D. (1987) Completion norms for 40 three-letter word stems, *Behavior Research Methods, Instruments, and Computers*, 19:422-445.
- Graf, S. (1952) *Measurement Of Hand Length, Muscular Control, and Motility Related To Handedness*. Unpublished master's thesis, Syracuse University.
- Gray, C. and Kopp, G. (1944) Voiceprint identification, *Report presented to the Bell Telephone Labs*, 1-14.
- Guttentag, R.E. (1985) Memory and aging: Implications for theories of memory development during childhood, *Developmental Review*, 5:56-82.
- Hagness, D.E. (1970) A preliminary investigation of the modified rhyme test as a test of speech discrimination, *Dissertation Abstracts International*, 31(2-B): 945-946.
- Hammersley, R. and Read, J.D. (1996) Voice identification by humans and computers, *Psychological issues in eyewitness identification* (S. L. Sporer, R.S. Malpass, G. Koehnken, eds.), Mahwah, NJ: Lawrence Erlbaum Associates: 117-152.
- Hamsher, K. (1990) Specialized neuropsychological assessment methods. *Handbook of psychological assessment* (2nd ed.). *Pergamon general psychology series*, (Gerald Goldstein, Michel Hersen, Eds.) 131: 256-279, New York: Pergamon Press, Inc.
- Handkins, R.E. and Cross, J.F. (1991) Voice similarity: its measurement and its effects on line-up fairness, Unpublished manuscript, St. Louis University.

- Hargreaves, D.J. and North, A.C. (1997) *The Social Psychology of Music*, Oxford, England: Oxford University Press.
- Hecker, M.H.L. (1971) Speaker recognition: An interpretive survey of the literature, ASHA, *Monograph 16*, Washington: DC.
- Hecker, M.H.L., Stevens, K., von Bismark, G. and Williams, C. (1968) Manifestations of task-induced stress in the acoustic speech signal, *Journal of the Acoustical Society of America*, 44:993-1001.
- Hicks, J.W., Jr. (1979) An acoustical/temporal analysis of emotional stress in speech, Ph.D. Dissertation, University of Florida.
- Hirano, M. (1981) *Clinical Examination of Voice*, New York: Springer-Verlag.
- Hirson, A. and Duckworth, M. (1995) Forensic implications of vocal creak as voice disguise, *BEIPHOL, Studies in Forensic Phonetics*, 64:67-76.
- Holden, R.H. (1984) Wechsler Memory Scale-Revised, *Test Critiques, VII* (eds. Keyser, D.J. and Sweetland, R.C), Kansas City: Test Corporation of America.
- Hollien, H. (1971) The peculiar case of "voiceprints", *Journal of the Acoustical Society of America*, 56:210-213.
- Hollien, H. (1977) Status report on "voiceprint" identification in the United States, *Proceed., International Conference of Crime Countermeasures*, Oxford, UK, 9-20.
- Hollien, H. (1990) *The acoustics of crime, the new science of forensic phonetics*, New York: Plenum Press.
- Hollien, H. (1992) Noisy tape recordings in forensics, *Proceeding Speech Processing in Adverse Conditions*, Cannes-Mandelieu, France, 167-170.
- Hollien, H. (1996) Consideration of guidelines for earwitness lineups, *Forensic Linguistics* 3:14-23.
- Hollien, H. (1997) Basic Criteria for Earwitness Lineups, draft for the International Association of Forensic Phonetics.
- Hollien, H., Bennett, G., and Gelfer, M.P. (1983) Criminal identification comparison: Aural versus visual identifications resulting from a simulated crime, *Journal of Forensic Sciences*, 28:208-221.

- Hollien, H., Childers, D.G. and Doherty, E.T. (1977) Semi-automatic speaker identification system (SAUSI), *Proceedings of the IEEE, ICASSP*, 26:768-771.
- Hollien, H., Gelfer, M.P. and Huntley, R. (1990a) The natural speech vector concept in speaker identification, *Neue Tendenzen in der Angewandten Phonetik III*, 71-87.
- Hollien, H., Hicks, J.W. and Oliver, L.H. (1990b) A semi-automatic system for speaker identification, *Neue Tendenzen in der Angewandten Phonetik III*, 89-106.
- Hollien, H., Huntley, R., Künzel, H., and Hollien, P.A. (1995) Criteria for earwitness lineups, *Forensic Linguistics* 2(2):143-53.
- Hollien, H. and Köster, J.P. (1996) Speaker identification by aural-perceptual approaches, *BEIPHOL*, *Miszellen IX*, 46: 135-152.
- Hollien, H. and Majewski, W. (1977) Speaker identification by Long-term spectra under normal and distorted speech condition, *Journal of the Acoustical Society of America*, 62:975-980.
- Hollien, H., Majewski, W. and Doherty, E.T. (1982) Perceptual identification of voices under normal, stress and disguised speaking conditions, *Journal of the Forensic Sciences*, 10:139-148.
- Hollien, H., Majewski, W. and Hollien, P.A. (1975) Analysis of F0 as a speaker identification technique, *Eighth Internat. Congress of Phonetic Sciences, Abstract of Papers*, 337.
- Hollien, H. and Martin, C.A. (1996) Conducting research on the effects of intoxication on speech, *Forensic Linguistics*, 3(1):107-128.
- Hollien, H. and McGlone, R.E. (1976) An evaluation of the "voiceprint" technique of speaker recognition, *Proceedings of the Carnahan Conference on Crime Countermeasures*, 30-45, (1976; reprinted in *Nat. J. Crim. Def.* 2:117-130, and in *Course Handbook*, The institute of Continuing Legal Education, Ann Arbor, Michigan, 391-404.
- Hollien, H. and Schwartz, R. (1997) Noncontemporary speech and speaker identification (submitted).
- Hollien, H. and Thompson, C.L. (1990) Effects of listening experience on decoding speech in HeO2 environments, *Diving for Science*, W.C. Jaap (ed.), 179-191.
- Horst, J.W. (1987) Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects, *Journal of the Acoustic Society of America*, 82(3):874-885.



- Hosmer, D.W. and Lemeshow, S. (1989) *Applied Logistic Regression*, New York: John Wiley & Sons.
- Houde, R.A. (1995) Pitch perception physiology and psychophysics as a basis for the design of pitch detection algorithms, *Journal of the Acoustical Society of America*, 97:3243.
- House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D. (1963) Psychoacoustic speech test: A modified rhyme test, *United States Air Force Electronic Systems Division Technical Documentary Report*, 63-403.
- House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D., (1965) Articulation testing methods: consonantal differentiation with a closed-response set, *Journal of the Acoustical Society of America*, 37: 158-166.
- Huntley, R. (1992) Listener skill in voice identification, paper presented at the American Academy of forensic Sciences, Boston, M.A., USA.
- Huntley, R. and Pass, K. (1995) Task influences on earwitness reliability, *BEIPHOL, Studies in Forensic Phonetics*, 64:121-131.
- Iles, M. (1972) Speaker identification as a function of fundamental frequency and resonant frequencies, Ph.D. Dissertation, University of Florida.
- Imaizumi, S., Mori K., Kiritani, S., Kawashima, R. Sugiura, M, Fukuda, H., Itoh, K., Kato, T., Nakamura, K., Hatano, K., Kojima, S. and Nakamura, K. (1997) Vocal identification of speaker and emotion activates different brain regions, *Neuroreport*, 8(12): 2809-2812.
- Ingemann, F. (1968) Identification of the speaker's sex from voiceless fricatives, *Journal of the Acoustical Society of America*, v44(4):1142-1144.
- Jacoby, L.L. (1983) Perceptual enhancement: persistent effects of experience, *Memory and Cognition*, 9:21-38.
- Jacoby, L.L, Toth, H.P., and Yonelinas, A.P. (1993) Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122: 139-154.
- Jacoby, L.L, Yonelinas, A.P. and Jennings, J.M. (in press) The relation between conscious and unconscious (automatic) influences: A declaration of independence, *Scientific Approaches to the Question of Consciousness*, J. Cohen and J.W. Schooler, editors, Hillsdale, NJ: Erlbaum.

- Jesteadt, W., Wier, C.C. and Green, D.M. (1977) Intensity discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America*, 61:169-177
- Jiang, M. (1995) *Experiments on a speaker identification system*, Ph.D. Dissertation University of Florida.
- Jiang, M. (1996) Fundamental frequency vector for a speaker identification system, *Forensic Linguistics*, 3:95-106.
- Johnson, C.C., Hollien, H. and Hicks, J.W., Jr. (1984) Speaker identification utilizing selected temporal speech features, *J. Phonetics*, 12:3(19-327).
- Joos, M.A. (1948) Acoustic phonetics, *Language* 24 (Supple.2): 1-136.
- Katz, J. (1994) *Handbook of Clinical Audiology*, Baltimore: Williams and Wilkins.
- Kavale, K. (1981) The relationship between auditory perceptual skills and reading ability: meta-analysis, *Journal of Learning Disabilities*, 14:539-546.
- Kellogg, R., Newcombe, C., Kammer, D., and Schmitt, K. (1996) Attention in direct and indirect memory tasks with short- and long-term probes, *American Journal of Psychology*. 109(2): 205-217.
- Kersta, L.G. (1962a) Voiceprint identification, *Journal of the Acoustical Society of America*, 34:725 (A).
- Kersta, L.G. (1962b) Voiceprint identification, *Nature*, 4861:1253-1257.
- Kimura, D. (1964) Left-right differences in perception of melodies, *Quarterly Journal of Experimental Psychology*, 16: 355-358.
- Kirsner, K., Dunn, J.C., and Standen, P. (1989) Domain-specific resources in word recognition, *Implicit Memory: Theoretical Issues*, (Eds. S. Lewandowsky, J. Dunn, and K. Kirsner), Hillsdale, New York: Erlbaum
- Koenig, B.E. (1986) Spectrographic voice identification: A forensic Survey, Letter to the Editor, *Journal of the Acoustical Society of America*, 79:2088-2090.
- Kogan, K.L. (1949) A review of the Wechsler Memory Scale. In O.K. Buros (Ed.), *The third Mental Measurements Yearbook*, 364-365, New Jersey: The Gryphon Press.
- Kosiel, U. (1973) Statistical analysis of speaker-dependent differences in the long-term average spectrum of Polish speech, *Speech Analysis and Synthesis* (W. Jassem, Ed.), Warsaw, Poland, 3:180-208.

- Köster, J.P. (1981) Auditive Sprechererkennung bei Experten und Naiven, *Festschrift Wängler*, Beiphol 52, 171-180.
- Köster, O., Hess, M.M, Schiller, N.O. and Künzel, H.J. (1997) The correlation between auditive speech sensibility and speaker recognition ability, presented at the International Association of Forensic Phonetics, Edinburgh, July 1997.
- Kreiman, J. and VanLancker, D. (1988) Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language*, 34(2): 246-252.
- Kreiner, D.S., Alvarado, M. and Shockley, S.R. (1997) Memory characteristics of college students who volunteer for research, *Psychological Reports*, 80(2) 515-520.
- Kucera, H. and Francis, W.N. (1967) *Computational analysis of present-day American English*, Providence, Rhode Island: Brown University Press.
- Künzel, H. (1990) *Phonetische Untersuchungen zur Sprechererkennung durch Linguistisch Naive Personen*, Stuttgart: Stienen-Verlag.
- Künzel, H. (1994) On the problem of speaker identification by victims and witnesses, *Forensic Linguistics*, 1:45-58.
- Kwalwasser, J. and Dykema, P.W. (1930) *Kwalwasser-Dykema Music Tests*, New York; Carl Fischer.
- Ladefoged, P. (1980) What are linguistic sounds made of? *Language*, 56:485-502.
- Ladefoged, P., and Vanderslice, R., (1967) The voiceprint mystique, *Working Papers in Phonetics* 7, University of California, Los Angeles, Nov.
- Lamp, C.J. and Keys, N. (1935) Can aptitude for specific musical instruments be predicted? *Journal of Educational Psychology*, 26: 587-596.
- LaRiviere, C.L (1971) Some acoustic and perceptual correlates of speaker identification, Ph.D. Dissertation, University of Florida.
- Laubstein, A.S. (1997) Problems of voice line-ups, *Forensic Linguistics* 4(2): 262-279.
- Laver, J. (1980) *The Phonetic Description of Voice Quality*, Cambridge: Cambridge University Press.
- Liberman, A.M. (1982) On finding that speech is special, *American Psychologist*, 37: 148-167.

- Lindsay, D.S. (1990) Misleading suggestions can impair eyewitnesses' ability to remember event details, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16:1077-1083.
- Linton, M. (1975) Memory for real-world events, In D.A. Norman and D.E. Rumelhart (Eds.), *Explorations in Cognition*, 376-404. San Francisco: Freeman.
- Linton, M. (1986) Ways of searching and the contents of memory. In D.C. Rubin (Ed.), *Autobiographical memory*, 50-67, Cambridge, England: Cambridge University Press.
- Loftus, E.F. (1975) Leading questions and the eyewitness report, *Cognitive Psychology*, 7:560-572.
- Loftus, E.F. (1977) Shifting human color memory. *Memory and Cognition*, 5:696-699.
- Loftus, E.F. (1979) *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.
- Loftus, E.F. (1980) *Memory*, Reading Massachusetts: Addison-Wesley.
- Loftus, E.F. (1986) Ten years in the life of an expert witness. *Law and Human Behavior*, 10(3): 241-263.
- Loftus, E.F. (1992) When a lie becomes memory's truth: Memory distortion after exposure to misinformation, *Current Directions in Psychological Science*, 1:121-123.
- Luck, J.E. (1969) Automatic speaker verification using cepstral measurements, *Journal of the Acoustical Society of America*, 46:1026-1032.
- Lumnis, R.C. (1972) Implementation of an on-line speaker verification scheme, *Journal of the Acoustical Society of America*, 52:S181.
- Lumnis, R.C. (1973) Speaker verification by computer using speech intensity for temporal registration, *IEEE Trans. Audio Electroacoust.*, AU-21:80-89.
- Mack vs. State of Florida (1907) 54 Fla. 55, 44 So. 706, citing 5, *Howell's State Trials*, 1186.
- Magoun, H.W. (1963) *The Waking Brain*. Springfield, Ill: Thomas.
- Majewski, W. and Hollien, H. (1974) Euclidean distance between Long-term speech spectra and a criterion for speaker identification, *Proceedings of the Speech and Communication Seminar-74*, Stockholm, Sweden, 3:303-310.

- Malpass, R.S. and Devine, P.G. (1983) Measuring the fairness of eyewitness identification lineups. *Evaluating witness evidence* (S. Lloyd-Bostock and B.R. Clifford, eds.). Chichester: Wiley.
- Mange, C. (1960) Relationships between selected auditory perceptual factors and articulatory ability, *Journal of Speech and hearing Research*, 3:367-374.
- Mann, V., Diamond, R., and Carey, S. (1979) Development of voice recognition: Parallels with face recognition, *Journal of Experimental Child Psychology*, 27:153-165.
- Mann, V.A. and Liberman, A.M. (1983) Some differences between phonetic and auditory modes of perception, *Cognition*, 14: 211-235.
- Markel, J.D., Oshika, B.T. and Gray, A.H. (1977) Long-term feature averaging for speaker recognition, *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-25:330-337.
- Masson, M.E.J., and Macleod, C.M. (1992) Reenacting the route to interpretation: Context dependency in encoding and retrieval, *Journal of Experimental Psychology: General*, 121:145-176.
- Masthoff, H. (1996) A report on a voice disguise experiment, *Forensic Linguistics, The International Journal of Speech, Language and the Law*, 3(1):160-167.
- Matarazzo, J.D. (1972) *Wechsler's measurement and appraisal of adult intelligence*, (5th ed.). New York: Oxford University Press.
- May, W.V. (1985) Musical style preferences and aural discrimination skills of primary grade school children, *Journal of Research in Music Education*. 33(1): 7-22.
- Mayes, A.R (1995) The assessment of memory disorders, *Handbook of memory disorders*. (Alan D. Baddeley, Barbara A. Wilson, Fraser N. Watts, Eds.) 367-391, England: John Wiley & Sons.
- McCandless, S.S. (1974) An algorithm for automatic formant extraction using linear prediction spectra, *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-22:135-141.
- McClelland, E.J. (1994) Phonetic and linguistic aspects of voice disguise in criminal proceedings, paper presented at the International Association of Forensic Phonetics Conference, Cardiff, U.K.

- McCloskey, M., and Zaragoza, M.S. (1985) Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses, *Journal of Experimental Psychology: General*, 114:1-16.
- McGehee, F. (1937) The reliability of the identification of the human voice, *Journal of General Psychology*, 17:249-271.
- McGehee, F. (1944) An experimental study of voice recognition, *Journal of General Psychology*, 31:53-65.
- McGlone, R., Hollien, H. and Hollien, P. (1977) Acoustic analysis of voice disguise related to voice identification, *Proceedings of the International Conference Crime Countermeasures*, 113:31-35.
- McKee, G. Humphrey, B. and McAdam, D.W. (1973) Scaled lateralization of alpha activity during linguistic and musical tasks. *Psychophysiology*, 10(4): 441-443.
- Mead, K.O. (1974) Identification of speakers from fundamental frequency contours in conversational speech, *Joint Speech Research Unit, Report 1002*.
- Mealey, L., Daood, C., and Krage, M. (1996) Enhanced memory for faces of cheaters, *Ethology and Sociobiology*, 17(2): 119-128.
- Meehan, E.F. and Pilotti, M. (1996) Auditory priming in an implicit memory task that emphasizes surface processing, *Psychonomic Bulletin and Review*, 3(4):495-498.
- Mehler, J., Bertoncini, J., Barrière, M., and Jassik-Gerschenfeld, D. (1978) Infant recognition of mother's voice, *Perception*, 7:491-497.
- Meltzer, D and Lehist, I. (1972) Vowel and speaker identification in natural and synthetic speech, *Journal of the Acoustical Society of America*, 51:S131.
- Mendenhall, W., Wackerly, D.D. and Scheaffer, R.L. (1990) *Mathematical Statistics With Applications*, Belmont California: Duxbury Press.
- Mensh, I.N. (1953) A review of the Wechsler Memory Scale. In O.K. Buros (Ed.), *The 4th Mental Measurements Yearbook*, 364-365, New Jersey: The Gryphon Press.
- Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information, *Psychological Review*, 63:81-97.
- Milner, B. (1962) Laterality effects in audition. *Interhemispheric Relations and Cerebral Dominance* (V.B. Mountcastle, ed.), Baltimore: The Johns Hopkins Press.

- Monroe, M. (1932) *Children who cannot read*, Chicago: University of Chicago Press.
- Monsen, R.B. and Engebretson, A.M. (1977) Study of variations in the male and female glottal wave, *Journal of the Acoustical Society of America*, 62:981-993.
- Moore, B.C.J. (1997) *An Introduction to the Psychology of Hearing*, 4th edition, New York: Academic Press.
- Moore, R. (1966) The relationship of intelligence to creativity. *Journal of Research in Music Education*, 14: 243-253.
- Morgan, A.H., McDonal, P.J. and MacDonald, H. (1971) Differences in bilateral alpha activity as a function of experimental task, with a note on lateral eye movements and hypnotizability. *Neuropsychologia*, 9: 459-469.
- Morrell, L.K. and Salamy, J.G. (1971) Hemispheric asymmetry of electrocortical responses to speed stimuli. *Science*, 174: 164-166.
- Mueller, J.H., Rankin, J.L. and Carlomusto, M. (1979) Adult age differences in free recall as a function of organization and method of presentation, *Journal of Gerontology*, 34:375-380.
- Mulligan, N.W. and Hartman, M. (1996) Divided attention and indirect memory tests, *Memory & Cognition*. 24(4): 453-465.
- Murdock, B. (1974) *Human Memory: Theory and Data*, Hillsdale, NY: Erlbaum.
- Nabelek, A.K. and Mason, D. (1981) Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms, *Journal of Speech and Hearing Research*, 24(3): 375-383.
- Naik, J.M. (1990) Speaker verification: a tutorial, *IEEE Communications Magazine*, 42-48.
- Nakasone, H. and Melvin, C. (1988) Computer assisted voice identification system (CAVIS), *Proceedings of the international Conference on Acoustics, Speech and Signal Processing*, 587-590.
- Nass, M.L. (1990) On hearing and inspiration in the composition of music, in *Psychoanalytic Explorations in Music* (S. Feder, R.L. Karmel, G.H. Pollock, eds.), Madison, CT: International Universities Press, Monograph 3: 179-193.
- Neisser, U. (1981) John Dean's memory: A case study, *Cognition*, 9:1-22.

- Nerbonne, G. (1968) The identification of speaker characteristics on the basis of aural cues, (Doctoral dissertation, University of Pennsylvania) *Dissertation Abstracts*, 28:10-B:4332-4333.
- Nixon, C.W., McKinley, R.L. and Moore, T.J. (1982) Increase in jammed word intelligibility due to training of listeners, *Aviation, Space, and Environmental Medicine*, 53(3): 239-244.
- Nolan, F. (1983) *The phonetic bases of speaker recognition*, Cambridge: Cambridge University Press.
- Nolan, F. and Grabe, E. (1996) Preparing a voice lineup, *Forensic Linguistics* 3(1):74-94.
- Ochsner, K.N., Chiu, C.Y.P., and Schacter, D.L. (1994) Varieties of priming, *Current Opinion in Neurobiology*, 4: 189-194.
- Ohman, S.E.G. (1966a) Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, 39:151-168.
- Ohman, S.E.G. (1966b) Perception of segments of VCCV utterances, *Journal of the Acoustical Society of America*, 40:979-988.
- O'Loughlin, B.J. and Moore, B.C.J. (1981) Improving psychoacoustical tuning curves, *Hearing Research*, 5:343-346.
- Orton, S.T. (1964) *Reading, writing, and speech problems in children*, New York: WW Norton and Company.
- Osborne, K. and Gale, O. (1976) Bilateral EEG differentiation of stimuli. *Biological Psychology*, 4: 185-196.
- Owen, Th. J. (1996) Voice identification, past and present. *Proceedings of the American Academy of Forensic Sciences Conference 1996*, Nashville, p57.
- Palmeri, T.J., Goldinger, S.D., and Pisoni, D.B. (1993) Episodic encoding of voice attributes and recognition memory for spoken words, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2):309-328.
- Papcun, G., Kreiman, J. and Davis, A. (1989) Long-term memory for unfamiliar voices, *Journal of the Acoustical Society of America*, 85:913-925.
- Patterson, R.D. (1976) Auditory filter shapes derived with noise stimuli, *The Journal of the Acoustical Society of America*, 59:640-654.



- Patterson, R.D., Nimmo-Smith, I., Weber, D.L. and Milroy, R. (1982) The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold, *Journal of the Acoustical Society of America*, 72:1788-1803.
- Peterson, G.E., and Barney, H.L. (1952) Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, 24:175-184.
- Pickles, J.O. (1988) *An Introduction to the Physiology of Hearing*, New York: Academic Press.
- Pisoni, D.B. (1993) Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning, *Speech Communication*, 13(1-2):109-125.
- Plomp, R. (1978) Auditory handicap of hearing impairment and the limited benefit of hearing aids, *Journal of the Acoustical Society of America*, 63(2):533-549.
- Plomp, R. (1986) A signal-to-noise ratio model for the speech reception threshold of the hearing impaired, *Journal of speech and hearing Research*, 29:146-154.
- Pollack, I., Pickett, J. and Sumby, W. (1954) On the identification of speakers by voice, *Journal of the Acoustical Society of America*, 26:403-406.
- Posner, M.I. and Keele, S.W. (1970) Retention of abstract ideas, *Journal of Experimental Psychology*, 83(2): 304-308.
- Potter, R.K. (1945) Visible patterns of speech, *Science*, 2654:463-470.
- Potter, R.K., Kopp, G.A., and Green, H.C. (1947) *Visible speech*, New York: D. Van Nostrand.
- Rajasekaran, P.K. (1984) Real-time factoring of the linear prediction polynomial of speech signals, *Digital Signal Processing '84*, (V. Cappallini and A.G. Constantinidis eds.), Elsevier Science Publishers, 405-410.
- Reich, A.R. and Duke, J.E. (1979) Effects of selected vocal disguise upon speaker identification by listening, *Journal of the Acoustical Society of America*, 66:1023-1028.
- Ritsma, R.J., Wit, H.P. and van der Lans, W.P. (1980) Relations between hearing loss, maximal word discrimination score and width of psychophysical tuning curves, *Psychophysical, Physiological and Behavioural Studies in Hearing* (eds. G. Van den Brink and F.A. Bilsen), Delft, The Netherlands: Delft University Press.

- Robbins, T.W. (1997) Arousal systems and attentional processes, *Biological psychology*, 45: 57-71.
- Roe, C.M. and Schwartz, M.F. (1996) Characteristics of previously forgotten memories of sexual abuse: A descriptive study, *Journal of Psychiatry & Law*, 24(2):189-206.
- Roediger, H.L. and McDermot, K.B. (1993) Implicit memory in normal human subjects, *Handbook of Neuropsychology* (Eds. F. Boller and J. Grafman), 8:63-131, Amsterdam: Elsevier.
- Rose, P. and Duncan, S. (1995) Naïve auditory identification and discrimination of similar voices by familiar listeners, *Forensic Linguistics, The International Journal of Speech Language and the Law*, 2(1):1-17.
- Rosen, S. and Fourcin, A. (1986) Frequency selectivity and the perception of speech, *Frequency Selectivity in Hearing*, London: Academic Press Inc. 373-487.
- Rosenberg, A.E. (1976) Automatic speaker verification: a review, *Proc. IEEE*, 64:475-487.
- Rosenberg, A.E. and Sambur, M.R. (1975) New techniques for automatic speaker verification, *IEEE Trans. Acoustics, Speech Signal Processing*, ASSP-23:169-176.
- Ross, E.D. (1981) The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere, *Archives of Neurology*, 38: 561-569.
- Rothman, H.B (1977) A perceptual (aural) and spectrographic identification of talkers with similar sounding voices, *Proceedings of the International Conference on Crime Counter measures*, Oxford, UK, 37-42.
- Rubin, D.C., Wetzler, S.E. and Nebes, R.D. (1986) Autobiographical memory across the lifespan. In D.C. Rubin (Ed.), *Autobiographical Memory*, 202-221. Cambridge, England: Cambridge University Press.
- Russell, E.W. (1975) A multiple scoring method for the assessment of complex memory functions, *Journal of Consulting and Clinical Psychology*, 43:800-809.
- Russell, E.W. (1981) The pathology and clinical examination of memory, *Handbook of Clinical Neuropsychology* (eds. S.B. Fiskow, and T.J. Boll), 287-319.

- Rusted, J.M. and Dighton, K. (1991) Selective processing of threat-related material by spider phobics in a prose recall task, *Cognition and Emotion*, 5(2): 123-132.
- Sachs, J.S. (1974) Memory in reading and listening to discourse, *Memory and Cognition*, 2:95-100.
- Sambur, M.R. (1975) Selection of acoustic features for speaker identification, *IEEE Trans. on Acoustics, Speech and Signal Process.* ASSP-24:283-287.
- Sambur, M.R. (1976) Speaker recognition using orthogonal linear prediction, *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-24:283-289.
- Saslove, H. and Yarmey, A. (1980) Long-term auditory memory: speaker identification, *Journal of Applied Psychology*, 45: 111-116.
- Sawyer, D.J. (1981) The relationship between selected auditory abilities and beginning reading achievement, *Language, Speech, Hearing, Service Schools*, 12:95-99.
- Schacter, D.L. (1990) Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. *Annals of the New York Academy of Sciences: Development and Neural Bases of Higher Cortical Function*, (Ed. A. Diamond), 608:543-571, New York: New York Academy of Sciences.
- Schacter, D.L., Church, B.A., and Osowiecki, D.M. (1994) Auditory priming in elderly adults: impairment of voice-specific implicit memory, *Memory*, 2(3):295-323.
- Scherer, I. (1974) Voice quality analysis of American and German speakers, *Journal of Psycholinguistic Research*, 3:281-298.
- Scherer, K.R. (1974) Acoustic concomitants of emotional dimensions: judging affect from synthesized tone sequences, *Nonverbal Communications*, (S. Weitz, Ed.), New York: Oxford University Press, 249-253.
- Scherer, K.R. (1977) Effects of stress on fundamental frequency of the voice, *Journal of the Acoustical Society of America*, 62:S25-26(A).
- Scherer, K.R. (1979a) Personality markers in speech, *Social Markers in Speech* (K.R. Scherer and H. Giles, Eds. ), Cambridge: Cambridge University Press, 147-209.
- Scherer, K.R. (1979b) Non-linguistic vocal indicators of emotion and psychopathology, *Emotions and psychopathology* (C.E. Izard, Ed.), New York, Plenum Press.
- Scherer, K.R. (1981) Vocal indicators of stress, *Speech Evaluation of Psychiatry* (J. Darby ed.), New York: Grune and Stratton.

- Scherer, K.R. (1986) Vocal affect expression: a review and a model for future research, *Psychological Bulletin*, 99(2): 143-165.
- Schiller, N.O. and Köster, O. (1996) Evaluation of a foreign speaker in forensic phonetics, a report. *Forensic Linguistics, The International Journal of Speech, Language and the Law*, 3(1):176-185.
- Schlaug, G., Jancke, L., Huang, Y., and Steinmetz, H. (1995) Musical ability, *Science*, 268(5211): 621-622.
- Schmajuk, N.A. and DiCarlo, J.J. (1991) Neural dynamics of hippocampal modulation of classical conditioning, *Neural Network Models of Conditioning and Action*, M.L. Commons, S. Grossberg, J. Staddon, editors, Hillsdale, NJ: Lawrence Erlbaum Associates, 149-180.
- Schmitter-Edgecombe, M. (1996) The effects of divided attention on implicit and explicit memory performance, *Journal of the International Neuropsychological Society*, 2(2): 111-125.
- Schwartz, M.F. (1968) Identification of speaker sex from isolated, voiceless fricatives, *Journal of the Acoustical Society of America*, 43(5):1178-1179.
- Schwartz, R. (1995) Effect of non-contemporary speech on aural-perceptual speaker identification, Conference of the International Association for Forensic Phonetics, Orlando, Florida.
- Seashore, C. Lewis, D., and Saezeit, J. (1960) *Manual-Seashore Measures of Musical Talents*, New York: The Psychological Corporation.
- Searleman, A. and Herrmann, D. (1994) *Memory From a Broader Perspective*, New York: McGraw-Hill, Inc.
- Sek, A. and Moore, B.C.J. (1995) Frequency discrimination as a function of frequency, measured in several ways, *Journal of the Acoustical Society of America*, 97: 2479-2486.
- Sergeant, D. (1973) Measurement of pitch discrimination, *Journal of Research in Music Education*, 21: 3-19.
- Shailer, M.J. and Moore, B.C.J. (1983) Gap detection as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America*, 74(2):467-473.
- Sheffert, S.M. and Fowler, C.A. (1995) The effects of voice and visible speaker change on memory for spoken words, *Journal of Memory and Language*, 34(5):665-685.

- Sher, A. and Owens, E. (1974) Consonant confusions associated with hearing loss above 2000 Hz., *Journal of Speech and Hearing Research*, 17:669-681.
- Sherbon, J.W. (1975) The association of hearing acuity, diplacusis, and discrimination with music performance, *Journal of Research in Music Education*, 23: 249-257.
- Shipley-Brown, F., Dingwall, W.O., Berlin, C.J., Yeni-Komshian, G. and Gordon-Salant, S. (1988) Hemispheric processing of affective and linguistic intonation contours in normal subject, *Brain and Language*, 33: 16-26.
- Shirt, M. (1983) An auditory speaker-recognition experiment comparing the performance of trained phoneticians and phonetically naive listeners, *Working Papers Linguistics and Phonetics*, 1: 115-117.
- Shirt, M. (1984) An auditory speaker recognition experiment, *Proceedings of the 1st Conference on Police Applications in Speech, Tape Recording Analysis*, Institute of Acoustics, London, 71-74.
- Silverman, F.H. and Silverman, E.M. (1975) Effects of threat of shock for being disfluent on fluency of normal speakers, *Percept. Motor Skills* 41:353-354.
- Simonov, P. and Frolov, M. (1973) Utilization of human voice for estimation of man's emotional stress and state of attention, *Aerospace Medicine*, 44: 256-258.
- Sivian, L.J. and White, S.D. (1933) On minimum audible fields, *Journal of the Acoustical Society of America*, 4:288-321.
- Soong, F.K., Rosenberg, A.E., Rabiner, L.R. and Juang, B.H. (1985) A vector quantization approach to speaker recognition, *International Conference on Acoustics, Speech and Signal Processing*, 387-390.
- Squire, L.R. (1987) *Memory and Brain*, New York: Oxford University Press.
- Stelmachowicz, P.G., Jestadt, W., Gorga, M.P. and Mott, J. (1985) Speech perception ability and psychophysical tuning curves in hearing-impaired listeners, *Journal of the Acoustical Society of America*, 77:620-627.
- Stevens, K.N., Williams, C.E., Carbonelli, J.R. and Woods, B. (1968) Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material, *Journal of the Acoustical Society of America*, 1596-1607.
- Stoval, J.V., Manning, W.H. and Shaw, C.K. (1977) Auditory assembly of children with mild and severe articulations, *Folia Phoniatrica*, 29:162-172.

- Stuntz, S.E. (1963) Speech intelligibility and talker recognition tests of Air Force communication systems. *Report. ESD-TDR-63-224*, Electronics Systems Division, Air Force Systems Command, Hanscom Field.
- Su, L.S., Li, K.P. and Fu, K.S. (1974) Identification of speakers by nasal coarticulation, *Journal of the Acoustical Society of America*, 156:1876-1882.
- Swain, C.E. (1972) The modified rhyme test: The effects of sensation level on discrimination scores on an adult sensorineural population, *Dissertation Abstracts International*, 32(10-B), 6105.
- Tartter, V.C. (1984) Laterality differences in speaker and consonant identification in dichotic listening, *Brain and Language* 23: 74-85.
- Tate, D. (1978) A study of speaker disguise, *unpublished Master's thesis*, University of Florida.
- Thompson, A.P. and LoBello, S.G. (1994) Reliable and abnormal scaled score range estimates for common Wechsler selected subtest short forms, *Journal of Psychoeducational Assessment*, 12(3) 264-269.
- Thompson, C. (1985a) Voice identification: speaker identifiability and correction of the record regarding sex effects, *Human Learning: Journal of Practical Research and Applications*, 4:(19-27.
- Thompson, C. (1985b) Voice identification: attempted recovery from a biased procedure, *Human Learning: Journal of Practical Research and Applications*, 4: 213-24.
- Thompson, C.P. (1987) A language effect in voice identification, *Applied Cognitive Psychology*, 1:121-131.
- Tilson, L.M. (1941) *The Tilson-Gretsch Musical Aptitude Test*. Chicago: Fred Gretsch.
- Tomatis, A. (1953) In R. Shuter, *The Psychology of Musical Abilities*, London: Methunen and Co., 1968.
- Tversky, B., and Tuchin M (1989) A reconciliation of the evidence on eyewitness testimony: Comments on McCloskey and Zaragoza. *Journal of Experimental Psychology: General*, 118:86-91.
- Tyler, R.S. (1979) Measuring hearing loss in the future, *British Journal of Audiology*, 13: 29-40.

- Tyler, R.S. (1986) Frequency resolution in hearing-impaired listeners, *Frequency selectivity in Hearing* (ed. B.C.J. Moore), London, Academic Press.
- Tyler, R.S., Summerfield, Q., Wood, E.J. and Fernandes, M.A. (1982) Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners, *Journal of the Acoustical Society of America*, 72(3):740-752.
- Van Lancker, D.R., Cummings, J.L., Kreiman, J., and Dobkin, B.H. (1988) Phonagnosia: A dissociation between familiar and unfamiliar voices, *Cortex*, 24:195-209.
- Van Lancker, D.R. and Kreiman, J. (1985) Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities, *Neuropsychologia*, 25: 829-834.
- Van Lancker, D.R., Kreiman, J., and Cummings, J. (1989) Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, 11:665-674.
- Velius, G. (1988) Variants of cepstrum based speaker identity verification, *Proceedings of the international Conference on Acoustics, Speech and Signal Processing*, 583-586.
- Viemeister, N.F. and Plack, C.J. (1993) Time analysis, *Human Psychophysics*, (eds. W.A. Yost, A.N. Popper and R.R. Fay), New York: Springer-Verlag.
- Wagenaar, W.A. (1988) *Identifying Ivan: A case study in legal psychology*, London: Harvester/Wheatsheaf.
- Wallin, N.L. (1991) *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perceptives on the Origins and Purposes of Music*, Stuyvesant, NY: Pendragon Press.
- Warnick, D.H. and Sanders, G.S. (1980) Why do eyewitnesses make so many mistakes ? *Journal of Applied Social Psychology*, 10:362-366.
- Webb, J.J. and Rissanen, E.L. (1993) Speaker identification experiments using HMM's, *International Conference on Acoustics, Speech and Signal Processing*, 2:387-390.
- Webster, J.C., Himes, H.W. and Lichtenstein, M. (1950a) San Diego County Fair hearing survey, *Journal of the Acoustical Society of America*, 22: 473-483.
- Webster, J.C., Lichtenstein, M. and Gales, R.S. (1950b) Individual differences in noise masked thresholds, *Journal of the Acoustical Society of America*, 22:483-490.

- Wechsler, D. (1945) A standardized memory scale for clinical use, *Journal of Psychology*, 19: 87-95.
- Wechsler, D. (1987) *A Manual for the Wechsler Memory Scale-Revised* (1987) San Antonio: The Psychological Corporation.
- Weir, C.C., Jesteadt, W., and Green, D.M. (1977) Frequency discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America*, 61:178-184.
- Welford, A.T. (1958) *Aging and Human Skill*, London: Oxford University Press.
- Wells, F.L. and Martin, H.A.A. (1923) A method of memory examination suitable for psychotic cases. *American Journal of Psychiatry*, 3: 243-258.
- Whellams, F.S. (1973) Musical abilities and sex differences in the analysis of aural-musical capacities. *Journal of Research in Music Education*, 21:30-39.
- Williams, C.E. (1964) The effects of selected factors on the aural identification of speakers, *Sect. III, Report ESD-TDR-65-153*, Electronics Systems Division, Air Force Systems Command, Hanscom Field.
- Williams, C.E., Hecker, M.H.L. and Kryter, K.D. (1965) Methods for psychoacoustic evaluation of speech communication systems, *United States Air Force Electronic systems Division Technical Documentary Report 65-153*.
- Williams, C.E. and Stevens, K.N. (1972) Emotions and speech: some acoustical correlates, *Journal of the Acoustical Society of America*, 52:1238-1250.
- Williams, T.F. (1982) A comparative study of three closed-response set speech discrimination tests, *Dissertation Abstracts International*, 42(7-B) 2767.
- Wing, H.D. (1948) *Standardized Tests of Musical Intelligence*, National Foundation for Educational Research.
- Wixted, J.T. and Ebbesen, E.B. (1991) On the form of forgetting, *Psychological Science*, 2:409-415.
- Wohlford, R.E. (1980) A comparison of four techniques for automatic speaker recognition, *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 908-911.
- Wolf, J., Krasner, M., Karnofsky, K., Schwartz, R., and Roucos, S. (1983) Further investigation of probabilistic methods for text-independent speaker identification, *International Conference on Acoustics, Speech and Signal Processing*, 551-554.



- Whybrow, P.C. (1994) Neuroendocrinology, *Biological Bases of Brain Function and Disease*, A. Frazer, P.B. Molinoff, and A. Winokur, editors, New York: Raven Press, 145-162.
- Yarmey, A.D. (1995) Earwitness speaker identification, *Special Issue: Witness Memory and Law, Psychology, Public Policy and Law*, 1(4):792-816.
- Yoshioka, P. and Thornton, A. (1980) Predicting speech discrimination from the audiometric thresholds, *Journal of Speech and Hearing Research*, 23:814-827.
- Young, A.W. (1984) Development of right hemisphere abilities. *Functions of the Right Cerebral Hemisphere* (A.W. Young ed.), London: Academic Press, 147-169.
- Zaidel, E. (1985) Language in the right hemisphere, *The Dual Brain: Hemispheric specialization in Humans*, D.F. Benson and E. Zaidel, editors, New York: Guilford Press: 205-231.
- Zalewski, J., Majewski, W. and Hollien, H. (1975) Cross-correlation between long-term speech spectra as a criterion for speaker identification, *Acustica*, 34:20-24.


## BIOGRAPHICAL SKETCH

Gea de Jong was born in Oudega, in The Netherlands, on February 12, 1968. She studied Computational Linguistics at the University of Groningen and specialized in artificial intelligence. In her fourth year, she developed a database management system for a study on the vocal characteristics of singers that was carried out at the Ear-Nose-Ear Clinic at the Academic Hospital in Groningen. As part of her study in Groningen, she spent her fifth year at the University of Cambridge, United Kingdom, as a visiting scholar. She studied the pronunciation of fricatives using electropalatography. In 1992, she obtained her Cum Laude Bachelor's degree of the University of Groningen. That year, she went again to Cambridge, now with the status of a master's student. She was admitted to Peterhouse, the oldest college of Cambridge, founded in 1254. She obtained her Master of Philosophy degree in "Computer Speech and Language Processing" in 1993. Her thesis was on using the temporal decomposition algorithm for automatic speaker verification.


Since arriving in Gainesville, Florida, in 1994 for her doctoral studies at the University of Florida, Gea has worked as a research assistant for Dr. Harry Hollien and as a teaching assistant at the Linguistics Department, the Speech Department and English Language Institute. She was also employed as a forensic assistant at Forensic Communications Associates and as a translator and database programmer. In her

dissertation she studies the effects of earwitness characteristics on speaker identification. The focus is on memory, hearing, and music. In 1997 she received the Outstanding Academic Achievement Award. Upon completing her Ph.D. in forensic phonetics, she will pursue her career further in Europe.


I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Harry Hollien, Chair  
Professor of Linguistics


I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
William Samuel Brown  
Professor of Communication  
Processes and Disorders

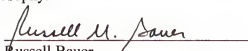
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Carl Crandell  
Associate Professor of  
Communication Processes and  
Disorders

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Anne Wyatt-Brown  
Associate Professor of Linguistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Russell Bauer  
Professor of Clinical and Health  
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

A handwritten signature in blue ink, consisting of a large, sweeping loop followed by a vertical line and a small flourish.

---

Antonius Broeders  
Director of Schrift en Spraak  
Onderzoek, Rijswijk, The  
Netherlands

This dissertation was submitted to the Graduate Faculty of the Department of Linguistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1998

---

Dean, Graduate School